AD NUMBER
ADB018239
LIMITATION CHANGES
TO: **Approved for public release; distribution is unlimited.**
FROM: **Distribution authorized to U.S. Gov't. agencies only; Administrative/Operational Use; APR 1977. Other requests shall be referred to Rome Air Development Center, Griffiss AFB, NY.**
AUTHORITY
RADC ltr 8 Nov 1982

THIS REPORT HAS BEEN DELIMITED AND CLEARED FOR PUBLIC RELEASE UNDER DOD DIRECTIVE 5200.20 AND NO RESTRICTIONS ARE IMPOSED UPON ITS USE AND DISCLOSURE.

DISTRIBUTION STATEMENT A

APPROVED FOR PUBLIC RELEASE;

DISTRIBUTION UNLIMITED.

RADC-TR-77-132
Final Technical Report
April 1977

SYSTEM AND MASS STORAGE STUDY FOR DEFENSE MAPPING AGENCY TOPOGRAPHIC CENTER (DMATC/HC)

Planning Research Corporation

Distribution limited to U. S. Gov't agencies only; test and evaluation; April 1977. Other requests for this document must be referred to RADC (ISFA), Griffiss AFB NY 13441.

ROME AIR DEVELOPMENT CENTER
AIR FORCE SYSTEMS COMMAND
GRIFFISS AIR FORCE BASE, NEW YORK 13441

This report has been reviewed and is approved for publication.

APPROVED: Yale Smith
YALE SMITH
Project Engineer

APPROVED: Robert D. Krutz
ROBERT D. KRUTZ, Colonel, USAF
Chief, Information Sciences Division

FOR THE COMMANDER: John P. Huss
JOHN P. HUSS
Acting Chief, Plans Office

Do not return this copy. Retain or destroy.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

19

1. REPORT NUMBER: 8 RADC-TR-77-132

2. GOVT ACCESSION NO.:

3. RECIPIENT'S CATALOG NUMBER: 9 14 PRC-R-2008

4. TITLE (and Subtitle): 6 SYSTEM AND MASS STORAGE STUDY FOR DEFENSE MAPPING AGENCY TOPOGRAPHIC CENTER (DMATC/HC).

5. TYPE OF REPORT & PERIOD COVERED: Final Technical Report, 28 Jun 76 — 31 Jan 77

6. PERFORMING ORG. REPORT NUMBER: R-2008

7. AUTHOR(s): 10 Richard K. Burkard, Robert J. McGrath, Brandon L. Buteau, Paul A. Scheffer, Nancy A. McDonald

8. CONTRACT OR GRANT NUMBER(s): 15 F30602-76-C-0338

9. PERFORMING ORGANIZATION NAME AND ADDRESS: Planning Research Corporation/PRC Information Sciences Co/7600 Old Springhouse Rd McLean VA 22101

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 63701B 17 03 16 32040310

11. CONTROLLING OFFICE NAME AND ADDRESS: Rome Air Development Center (ISFA) Griffiss AFB NY 13441

12. REPORT DATE: 11 Apr[redacted]977

13. NUMBER OF PAGES: 182

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): Same 12 179 p.

15. SECURITY CLASS. (of this report): UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE: N/A

16. DISTRIBUTION STATEMENT (of this Report)

Distribution limited to U. S. Gov't agencies only; test and evaluation; April 1977. Other requests for this document must be referred to RADC (ISFA), Griffiss AFB NY 13441.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

Same

18. SUPPLEMENTARY NOTES

RADC Project Engineer: Yale Smith (ISFA)

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Computer Architecture
Mass Storage Systems
Back-end Data Base Management

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The purpose of this study was to translate processing and data storage requirements of the Defense Mapping Agency Topographic Center for the FY77-FY83 time period to a system architecture capable of supporting the Center's mission. Requirements provided by the DMATC indicated that a processing capacity twice as large as FY76 workload would require an immediate augmentation. Projected data holdings indicated a requirement for storage of up to 4.7 x 10^{12} bits by FY83.

10 to the 12th power

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED 409 126

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

An advanced system concept which would accommodate an associative array processor, a mass storage system, and a back-end data base management system was developed. The system architecture is a form of distributed processing in which computer interties to a bus are accommodated through microprocessor ports. Data base management functions would be provided by a staging computer supported by disk storage which is shared and thus accessible by both intertied processors and a mass storage device. The concept would allow retention of large mainframes while the focal point of the information flow would be the staging center to which all transit files are passed. Key to the concept is a microprocessor interface device to control access between system components and the bus. Security constraints are a major factor in achieving the level of system integration needed by the Center. The concept is therefore dependent on the premise that advancing technology will allow the development of microprocessor devices and techniques to control access to data files in an acceptable manner.

Interim actions recommended to improve production capabilities included implementation of decentralized interactive edit capabilities supporting cartographic and photogrammetric production, and a program to upgrade data files through consolidation and possible use of a data base management system. Further analysis of data holdings and associated access and response requirements is recommended to provide a better basis for definition of a mass storage media.

ADDENDUM

Following submission of this report, the Planning Research Corporation was informed that a change in DMATC program direction was in process which will significantly modify workload projections as shown in this report. This change of program direction will involve increased emphasis on the production of digital topographic elevation data in matrix format. This change of direction is expected to cause a shift in the application of production resources and can be expected to affect the projected growth of pertinent data files as cited in this report. While the full effect of these changes is not apparent at this time, they do not appear to significantly alter the major conclusions and recommendations of this study.

ACKNOWLEDGEMENTS

The PRC Information Sciences Company study team wishes to express its appreciation to personnel of the DMA Topographic Center for assistance provided during the study. A true spirit of cooperation was exhibited toward us at every level of the organization. This cooperation was evidenced during the initial briefing by Center personnel and extended throughout the study period. The list of individuals who contributed to the effort would be extensive and represent personnel at all levels of the organization. We would, however, like to express our appreciation to the Commander, Colonel William R. Cordova, for his personal interest and encouragement; to Mr. Arthur Noma, our principal point of contact; and to Mr. Richard Vitek for valuable assistance in collecting data in the Center.

We would also like to cite the cooperation of the Cullinane Corporation, and specifically Mr. John Cullinane for providing documentation for use in the study reflecting Cullinane supported internal research.

TECHNICAL REPORT SUMMARY

Technical Problem. The purpose of this effort was to provide a study of advanced computer technology for the processing and mass storage of digital data at the Defense Mapping Agency Topographic Center (DMATC) and the application of this technology to provide more efficient operation and support of the digital data bases required by the DMATC in the FY 77 through FY 83 time period. The study was to be based on information processing requirements provided by the DMATC for translation into a statement of computer system architecture which would support the Center's production requirements.

General Methodology. The study was performed using three basic sources of information. The first consisted of information provided by the DMATC. An initial briefing was provided which described the environment and initial assessments of projected data processing and data storage requirements. Using this information, extensive interviews and briefings were held with data processing personnel, data file managers, and other production and management personnel. Selected documentation was provided along with access to computer job statistics and other forms of documentation. Information was also provided regarding current programs to increase production capabilities. Questions arising during the study were resolved with Center personnel through further interviews.

The second source of information consisted of briefings and interviews provided by research and development program managers at the U.S. Army Engineer Topographic Laboratories, Fort Belvoir, Virginia, and the Rome Air Development Center. These briefings provided further information relevant to the changing cartographic environment.

The third source of information consisted of a literature search and direct queries to other commercial organizations. These queries were concentrated on mass storage technology and emerging technology in back-end data base management systems.

Three briefings summarizing project status were provided the DMATC during the study effort.

Technical Results. The study effort concluded that achievement of long-range objectives would require parallel actions addressing processing capabilities as well as improved methods of information management. Study findings, based on DMATC provided processing requirements, support a major augmentation of the central processing capabilities. These requirements reflect a 100 percent increase during FY 77 with further growth beyond 1978 to nearly triple the 1976 level of usage. The projected requirements require augmented capabilities for both collateral and SAA processing. While system augmentation is essential to satisfy basic data processing requirements, other actions will be required to reduce production calendar time which is currently constrained by the existing batch mode of operations.

The most significant growth in data processing requirements and in the growth of digital data files is associated with the production of 1:50,000 scale topographic line maps. Automated cartography will play a major role in meeting significantly increased production work loads of line maps. (Note: Following completion of this report, a change in DMATC program direction occurred which will increase the production of digital terrain data. While this change will alter some of the projections cited in this report, it is not considered to alter the major conclusions reached in this study.)

The automated portion of this production process is supported by a software system designed to operate in a batch mode on the UNIVAC 1108 processor. The creation of digital files (digitizing, plotting, editing, and verification) requires a sequential series of operations on the UNIVAC 1108, which may require iteration until an error-free record is created. This process involves a high number of passes through the central processor, extending the time to create error-free digital records, supports a high level of tape activity, and increases the probability of human errors during job submission and data movement. Implementation of an interactive cartographic edit system to support the creation of master digital records is recommended. Consideration of a similar system to provide an editing and formatting for photogrammetric data is also recommended.

From an information management point of view, the current environment is characterized by a large number of independent functionally oriented fact files. Many of these files are interrelated functionally, but lack automated interfaces. Exploitation of these files to support planning and control functions is consequently time consuming. A file analysis program to examine restructuring of the files to more effectively support management functions is recommended.

Projected growth of data files and particularly those related to automated cartography indicate a potential data storage requirement of 4.7×10^{12} bits by 1983. These files constitute a source data base which could be used to support new production. Within the context of the study, the files are viewed as primarily archival in nature. Continuous refinement of this projection, including access frequency and response requirements is recommended to provide a better basis for selection of a mass storage media.

An advanced system architecture concept which would accommodate an associative array processor, a mass storage system, and a back-end data base management system was developed. While each of these areas is in advanced stages of development, full commitment at this time does not appear justifiable. Consequently, a system concept which would allow modular growth as technology and production requirements change is recommended.

The system architecture recommended is a form of distributed processing in which computer interties to a bus are accommodated through microprocessor ports. Under the concept data base management functions would be provided by a staging computer supported by disk storage which is shared and thus accessible by both intertied processors and a mass storage device. The principal motivation for the concept is a need to provide improved information transfer. The concept would allow retention of the extensive capacity of large main-

frames while the focal point of the information flow would be the staging center to which all transit files are passed. The staging center together with a system-wide communications bus allow all components in the environment to be tied together. Key to the concept is the microprocessor device which would control access to and between system components and the bus.

Security constraints are recognized as a major factor in achieving the level of system integration needed by the Center. The concept is therefore dependent on the premise that advancing technology will allow the development of techniques which will serve to control access to data files in an acceptable manner.

Implications for Further Research. The report identifies several tasks which are recommended for action by the DMATC which are within current state-of-the-art. The areas identified below are considered to be the most critical from a long-term research viewpoint.

- Research efforts directed toward the development and demonstration of microprocessor based devices and techniques to support a security control function should be given highest priority.

- Continued development of the back-end data base management concept is required to clearly establish the viability of this concept for the management of very large data bases.

- The development of significantly improved data processing capabilities supporting raster scan operations is urgently needed. A study effort examining all current capabilities should be initiated to identify both short and long term alternatives.

- Continued research to develop cost-effective options for mass storage is required. Prospects for affordable very large random access storage devices are dim.

- A fresh approach to examining the use of parallel processing techniques in automated cartography should be considered.

EVALUATION

1. The System and Mass Storage Study for the Defense Mapping Agency Topographic Center (DMATC/HC) was the second study performed by the PRC Information Sciences Company wherein the purpose was the translation of data processing and storage requirements into computer system architectures.

2. The first effort, which was also very well received by DMA, addressed the problem at DMA Aerospace Center. The present DMATC/HC study provided the user with extremely useful information at a time when DMA data processing requirements are becoming critical operational support areas. Recommendations for immediate subsystem improvements and computer system architecture are playing a key role in defining the computer upgrading being initiated at DMAAC and DMATC.

3. The study examined areas such as: alternative computer architectures, distributed processing and networking considerations (including multilevel security implications), interactive processing, mass storage and retrieval systems, back-end data management systems and software development and maintenance. All of these areas are key items within RADC technical program objectives.

Yale Smith

YALE SMITH
Project Engineer

CON. TS

		Page
1	BACKGROUND	1
1.1	Purpose and Goals	1
1.2	DMATC Organization and Mission	1
1.3	Mission Objectives Influencing ADP Operations	2
1.4	Organizational Factors	3
2	REQUIREMENTS ANALYSIS	5
2.1	Problem Formulation	5
2.1.1	Concentration of Study Areas	5
2.1.2	Study Approach	6
2.1.3	Organization of the Study	6
2.2	Analysis	6
2.2.1	Data Processing Environment	6
2.2.1.1	UNIVAC 1108 Central Processors	6
2.2.1.1.1	Site I	8
2.2.1.1.2	Site II	8
2.2.1.1.3	System Operation	8
2.2.1.2	Dedicated Input/Output Systems	8
2.2.1.3	Burroughs B-3500	9
2.2.1.4	IBM 7090/7094	10
2.2.2	Central System Utilization	11
2.2.2.1	Processing Activity	11
2.2.2.1.1	Disk Utilization	12
2.2.2.1.2	Evaluation of SIP Statistics	12
2.2.2.2	System Utilization by Program Area	13
2.2.3	Data Systems	14
2.2.3.1	Topographic Data Library System (TDLS)	17
2.2.3.2	Semi-Automatic Cartographic System (SACARTS)	18
2.2.3.3	Data Reduction Interface and Control System (DARIC)	18
2.2.3.4	DMA Program Management Information System (DMIS/P)	19
2.2.3.4.1	DMIS	19
2.2.3.4.2	DMIS/P	20
2.2.3.4.3	The DMATC DMIS/P	21
2.2.3.5	Digital Topographic Information Bank (DTIB)	21
2.2.3.6	Product Maintenance System (PMS) and Area Requirements and Production System (ARAPS)	23
2.2.3.7	Automated Map Research Report (AMRR)	24
2.2.3.8	CELESTE	25
2.2.3.9	Data System Summary	26
2.2.4	Current Data Storage	29
2.2.4.1	Magnetic Tapes	29
2.2.4.2	Disk Storage	29
2.2.5	Projected Data Processing Requirements	30
2.2.5.1	Overall Processing Requirement	30
2.2.5.2	Collateral Processing Requirement	32
2.2.5.3	SAA Processing Requirement	32
2.2.5.4	Implications of Projected Processing Work Load	32

CONTENTS (Continued)

		Page
2.2.5.5	Processing Work Load by Functional Areas	35
2.2.6	Projected Data Storage Requirements	35
2.2.6.1	Projected Source Data Base Holdings	38
2.2.6.2	Online Storage Requirements	38
2.2.7	Projected Restructuring of Data Files	40
2.2.7.1	Initial File Assessment	42
2.2.7.2	Interaction with Standards	43
2.2.8	Long-Range Projections--Pilot Digital Operations Plan	43
2.2.8.1	Systems Concept and Assumptions	43
2.2.8.2	Relationship to the System Architecture Study	44
3	CURRENT RESEARCH AND DEVELOPMENT PROGRAM	47
3.1	Raster Scanner	47
3.1.1	System Functions	47
3.1.2	Data Processing Impact	47
3.1.3	Impact on SACARTS Production	48
3.1.4	Consideration of Alternatives	49
3.1.5	Potential Use of the ETL STARAN Processor	50
3.1.6	Conclusions	50
3.2	Integrated Photogrammetric Instrumentation Network (IPIN)	51
3.2.1	Applicability of the IPIN to DMATC	54
3.2.2	Summary	55
3.3	Associative Array Processor	55
3.3.1	Definition of Terms	55
3.3.2	General Applications of AAP	58
3.3.3	STARAN Characteristics	58
3.3.4	DMATC Applications for the STARAN	61
3.3.4.1	Synopsis of Associative Array Processing of Raster Scanned Data for Automated Cartography	61
3.3.4.2	Synopsis of Application of a Parallel Processing Computer in LACIE	62
3.3.4.3	Synopsis of an Implementation of a Data Management System on a Associative Processor	62
3.3.4.4	Other Research and Development	63
3.3.4.4.1	Mass Storage Device	63
3.3.4.4.2	High Order Language Development	63
3.3.4.4.3	UNIVAC STARAN Interface Study	64
3.3.5	Recommendations Concerning AAP	64
4	ADVANCED INFORMATION HANDLING TECHNOLOGY	66
4.1	Back-End Data Base Management	66
4.1.1	Introduction	66
4.1.2	Definitions and Concepts	67
4.1.3	System Impact of Back-End Processing	69
4.1.4	Requirements for Back-End Processing	75
4.1.5	Status of Back-End Processing	76
4.2	Mass Storage Systems	77
4.2.1	Ampex Terabit Memory (TBM) System	77

CONTENTS (Continued)

		Page
4.2.1.1	Characteristics	77
4.2.1.2	Capacity and I/O Rates	79
4.2.1.3	Reliability	80
4.2.1.4	Software Impact	80
4.2.1.5	Cost and Status	80
4.2.2	CALCOMP 7110 Automated Tape Library	80
4.2.2.1	Characteristics	80
4.2.2.2	Capacity and I/O Rate	81
4.2.2.3	Reliability	81
4.2.2.4	Software Impact	81
4.2.2.5	Cost	81
4.2.3	Control Data 38500	81
4.2.3.1	Characteristics	82
4.2.3.2	Capacity and I/O Rates	82
4.2.3.3	Reliability	85
4.2.3.4	Software Impact	85
4.2.3.5	Cost	85
4.2.4	MASSTAPE	85
4.2.4.1	Characteristics	85
4.2.4.2	Capacity and I/O Rate	85
4.2.4.3	Reliability	86
4.2.4.4	Software Impact	86
4.2.4.5	Cost	86
4.2.5	IBM 3850	86
4.2.5.1	Characteristics	86
4.2.5.2	Capacity and I/O Rates	87
4.2.5.3	Reliability	87
4.2.5.4	Software Impact	87
4.2.5.5	Cost	90
4.2.6	Precision Instruments System 190	90
4.2.6.1	Characteristics	90
4.2.6.2	Capacity and I/O Rate	90
4.2.6.3	Reliability	90
4.2.6.4	Software Impact	91
4.2.6.5	Cost	91
4.2.7	The Human Readable/Machine Readable (HRMR) System	91
4.2.7.1	Research Prototype HRMR Microfilm Memory	92
4.2.7.2	Description of HRMR Engineering Model Microfilm Mass Memory	92
4.2.7.3	Summary of HRMR Cartographic Applications Study	95
4.2.7.4	Summary	98
4.2.8	Future Mass Storage Technology	99
4.2.8.1	Magnetic Tape	99
4.2.8.2	Disks (Magnetic and Magneto-Optic) and Drums	99
4.2.8.3	Optical	100
4.2.8.4	Solid State Storage	100

CONTENTS (Continued)

		Page
4.3	Security Issues	100
5	ARCHITECTURAL ALTERNATIVES	106
5.1	Factors Influencing System Architecture	106
5.1.1	Computer Work Load Requirements	106
5.1.2	Need for Work Load Distribution	106
5.1.3	Growth of Data Holdings	107
5.1.4	Improved Information Management	107
5.1.5	Automated Chart Production	107
5.1.6	Security	108
5.1.7	Associative Array Processing	108
5.1.8	Objectives of Pilot Digital Operations	108
5.2	Central Processor Augmentation	108
5.2.1	Projected Work Load	109
5.2.2	Central Processor	109
5.2.3	Increased Information File Transits	109
5.2.4	Security Constraints	110
5.3	Proposed System Architecture	111
5.3.1	Justification and Assumptions	113
5.3.2	The Intertie Concept	114
5.3.3	The Staging Computer	117
5.3.4	Additional Technical Characteristics	118
5.3.5	Security Considerations	120
5.3.5.1	Logical Security	120
5.3.5.2	Physical Security	121
5.3.6	Software Considerations	121
5.3.6.1	Port Software	121
5.3.6.2	Other Software Considerations	121
5.3.6.2.1	Mass Storage Systems	122
5.3.6.2.2	Associative Array Processors	122
5.3.6.2.3	Microprocessors	123
5.3.6.2.4	Applications Programs	123
5.3.7	System Comparison	123
5.3.7.1	Existing Methods and Procedures	123
5.3.7.2	Proposed Methods and Procedures	124
5.3.7.3	Scenarios for Job Flow Through New Configuration	125
5.3.8	Configuration Options	130
5.4	Implementation Sequence	130
5.5	Applicability of the Intertie Concept to DMAAC	134
6	CONCLUSIONS	135
6.1	Central Processing	135
6.2	Dedicated Processing Subsystems	135
6.3	Data Systems	136
6.4	Information Management	136
6.5	Security	136
6.6	Advanced Technology	136
6.6.1	Mass Storage Systems	137

CONTENTS (Continued)

		Page
6.6.2	Associative Array Processors	137
6.6.3	Microprocessor Technology	137
6.6.4	Back-End Data Base Management	137
7	RECOMMENDATIONS	138
A	SACARTS	140
A.1	Introduction	140
A.2	Background	140
A.3	SACARTS Development	141
A.4	Identification of Problem Areas	141
A.4.1	GIST Batch Mode Operation	141
A.4.2	Data Base Access	143
A.4.3	Data Capture Systems	146
A.4.3.1	Digital Planimetric Compiler	146
A.4.3.2	Comp-U-Grid and CALMA	148
A.5	SACARTS Expansion	149
A.5.1	Raster Scanner	149
A.5.2	Raster Plotter	150
A.5.3	Digital I/O Display Equipment	150
A.6	Summary	150
A.6.1	SACARTS Recap	151
A.6.2	SACARTS Conclusions	152
A.6.3	Recommendations	153
B	INTERACTIVE EDIT STATION FUNCTIONAL DESCRIPTION	154
B.1	Smart Work Station	154
B.2	Desirable Work Station Functions	154
B.3	Work Station Hardware	156
B.4	Communication Interface	157
B.5	Mainframe CPU	157
B.6	Background Functions	158
B.7	Mainframe CPU Hardware	158
	BIBLOGRAPHY	160

FIGURES

		Page
1.	Data Processing Environment	7
2.	Utilization of U-1108 Resources by Program Area FY 76	16
3.	Total Projected Requirement (Hours of U-1108 CPU Time)	31
4.	Projected Requirement--Collateral Work Load (Hours of U-1108 CPU Time)	33
5.	Projected Requirement--SAA Work Load (Hours of U-1108 CPU Time)	34
6.	IPIN System Configuration	52
7.	Associative Address Operation	57
8.	STARAN System	60
9.	Hardware and Software Essentials of Back-End Processing	68
10a.	Dedicated Multiprocessing BEP	70
10b.	Distributed Multiprocessing BEP	70
11.	Partially Freestanding BEP	71
12.	BEP as a Staged Storage/MSS Controller	73
13.	Transverse Scan Rotating Head and Tape	78
14.	Read/Write Head Configuration and Tape Format for CDC 38500	83
15.	CDC 38500 Mass Storage System Block Diagram	84
16.	Helical-Scan Rotor with Head	88
17.	Data Flow Controlled by 3830 Microprocessor	89
18.	HRMR Engineering Model Microfilm Mass Memory Block Diagram	94
19.	Capabilities Access Matrix	102
20.	Network Security	104
21.	Proposed Advanced Configuration	112
22.	Microprocessor Unit Functional Flow	116
23.	Possible Job Flow Scenarios for Application Programming	126
24.	Job Flow--Data Movement to a Satellite	127
25.	Job Flow--Data Base Update	128
26.	Job Flow--Balancing Loading Assignments	129
27.	Implementation Sequence	131
A-1.	SACARTS Functional Flow	142
A-2.	Example of an Alternative Retrieval Method	145
B-1.	A Typical Cartographic System	155

TABLES

		Page
1.	UNIVAC 1108 CPU Utilization (FY 76)	15
2.	Data Base Characteristics	27
3.	Additional Data Files and Characteristics	28
4.	Significant Changes in UNIVAC 1108 CPU Utilization (FY 76 to FY 77)	36
5.	Major Production Work Load (UNIVAC 1108 CPU Hours)	37
6.	Data Base Growth Projection	37
7.	Projected Magnetic Tape Holdings	39
8.	Implied Technical Characteristics of Pilot Digital Operations Subsystems	45
9.	HRMR Engineering Model Microfilm Mass Memory Performance Goals and Parameters	96

1. BACKGROUND

1.1 Purpose and Goals. This study was initiated on June 28, 1976 with the objective of translating the data processing requirements of the Defense Mapping Agency Topographic Center (DMATC) into an advanced systems architecture which could best utilize advanced computer science technology for the processing and storage of digital information. The potential use of advanced technology to support the production and maintenance of large digital data bases through the 1983 time period was a principal consideration in the study. To this end topics of special interest included mass storage technology, associative array processing, and an emerging concept of information management, i.e., back-end data base management. Additional areas of interest are the implications of two new applications of data processing for cartographic purposes, namely the raster scanner plotter and a photogrammetric post-processor.

The basic requirements toward which the advanced system architecture is addressed consist of the requirements for data processing and data maintenance as provided by the DMATC. The objective has been to define a system architecture which would allow system growth in a modular fashion and provide options for growth which could be accommodated within the advanced concept as they become available. An overall constraint throughout the study has been to provide an approach which would always assure the basic capacity to meet DMATC mission requirements. Once an overall advanced concept is established, specific solutions to problem areas can be isolated, defined, and implemented within the total conceptual framework.

After a background discussion, this report addresses the baseline requirements for the study. Subsequent sections discuss aspects of advanced technology as they may contribute to the satisfaction of these requirements and lead to the definition of an advanced architectural concept viewed by the project team as a viable long-term objective.

1.2 DMATC Organization and Mission. The Defense Mapping Agency Topographic Center was established on July 1, 1972 as a major component of the Defense Mapping Agency (DMA). With the establishment of the DMA, mapping, charting, and geodesy (MC&G) resources in the military departments were consolidated into a separate Defense agency. Prior to the establishment of the DMA, the Topographic Center was an agency of the U.S. Army known as the U.S. Army Topographic Command and in early times, the Army Map Service. The headquarters of the DMATC is located in Brookmount, Maryland.

The Topographic Center has as its primary mission the production and distribution of MC&G products to satisfy the operational requirements of the land combat forces in the Department of Defense. These products and services consist of a broad spectrum of MC&G end-items including maps, geodetic data, and related information. To support this mission, DMATC maintains its primary production facility collocated with the headquarters in the Washington, D.C., area and four field production facilities in San Antonio, Texas; Providence, Rhode Island; Kansas City, Missouri; and Louisville, Kentucky.

In August 1976, the Geodetic Survey Squadron (GSS) located at Cheyenne, Wyoming was transferred to the DMATC in a consolidation of field survey resources.

DMATC also operates map storage and distribution facilities for all DMA products in depots located in Philadelphia, Pennsylvania and Clearfield, Utah.

To fulfill its mission of providing MC&G support for the land combat forces of the DOD, DMATC provides the following services:

- Production and distribution of maps and geodetic data
- Collection, processing, and storage of topographic data on a worldwide basis
- Technical support and guidance to U.S. Army and U.S. Marine Corps topographic units
- Direct support for land combat weapon systems and the development of new topographic products to support weapon systems
- Specialized topographic products, studies, and reports

The Center also provides service to the Department of Defense through the maintenance of data files of topographic maps and geodetic data on a worldwide basis. The Center also provides ADP support and assistance to the DMA Headquarters and performs selected functions such as distribution and source collection requirements for the entire DMA structure.

1.3 Mission Objectives Influencing ADP Operations. The DMATC, and the entire DMA structure, is a service activity dedicated to support the operational forces of the Department of Defense. Its effectiveness is consequently measured in terms of the timeliness, quality, and quantity of its response to operational requirements. Because of this supportive responsibility, the production capacity of the DMATC must be characterized by responsiveness, a high level of quality control, capacity, and flexibility. At the same time, economy of resources both in terms of equipment and manpower is a continued constraint. While these factors may appear obvious, they have direct implications to the level of ADP support required by the Center in the performance of its mission. Consequently, they are fundamental considerations in addressing the data processing resources that will be needed by the Center over the next seven years. The study will therefore consider these characteristics as they will influence the advanced system architecture:

- Responsiveness. Organization and utilization of ADP resources must support timely reaction to priority requirements. While timeliness is relative, growing data banks of processed topographic data together with advances in communications will significantly improve capabilities to provide responsive support to land combat forces. While such a consideration may provide a long-term growth objective, in the current world it still implies the capability to respond to low-volume, high-priority requirements in a minimum time span.
- Quantity. Requirements for data may represent completely processed data, such as maps, or data extracted from data banks. While many requirements can be satisfied by normal production scheduling, the

ADP support system should be adaptive to rapid retrieval of virtually any quantity of data. The effective utilization of ADP technology can also contribute to maintenance of an optimal depot inventory level so long as a responsive capability to respond to urgent high volume requirements can be satisfied.

- Quality. Although cartographic production standards demand a high level of quality control, they do not in general present unusual requirements on the ADP support system. Quality control is, however, an essential consideration in the effective use of ADP resources. Without effective quality control during the production process, maximum ADP effectiveness will not be reached and production throughput will be adversely impacted both in quantity and time.

- Flexibility. The ability to respond to changing user requirements either in terms of the type of product or urgency requires adaptable ADP resources. Flexibility is frequently addressed from a viewpoint of adjusting work loads to accommodate new or changed product requirements. From a long-term view, flexibility largely depends on the content, format, and organization of data stored in a resource data bank. Data management then becomes an integral factor in system architecture.

- Economy of Resources. Limitations of resources both in investment and manpower are recognized constraints. Full utilization of resources, elimination of redundant data, preservation of useful processed data, and system improvements based on labor/investment trade-offs all contribute to long-term resource management. Achievement of long term resource management objectives will largely depend on how well raw and processed data is managed as a production resource. Again, the manner in which data is managed--how it is organized, stored, retrieved, and applied--is a fundamental consideration in developing an ADP architectural concept which best utilizes available resources.

1.4 Organizational Factors. Organization of the DMATC is pertinent to this study only to the extent that it contributes to an understanding of the data processing functions and the management of data files (including their generation, maintenance, and utilization of ADP resources.) The principal elements of the organization pertinent to this study are:

- Production Departments: Generation, maintenance, and exploitation of data files
  - Department of Technical Services
    - Operation of map and other resource files
  - Department of Geodesy
    - Generation of geodetic control including analytical aerotriangulation
    - Management of field survey activities
    - Operation of the Geodetic Survey Squadron ADP resources
    - Management of the Satellite Geophysics Programs
  - Department of Cartography
    - Production of standard and special map products

- Department of Field Offices
    - Support of compilation and color separation phases of map production
- Department of Distribution
    - Management of product distribution facilities and resources

- Central Computer Support: Operation of centralized computer facilities
    - Operation of two UNIVAC 1108 processors
    - Programming support for other production departments
    - Operation of a Burroughs B-3500 processor

- Staff Offices
    - Directorate of Programs Production and Operations (PPO)--responsible for production scheduling and resource allocation. Relatively new functions established within this Directorate include a Data Base Administrator overseeing the management of data resources and a Quality Control Function.
    - Directorate of Plans Requirements and Technology (PRT)--responsible for assuring the technical capacity of DMATC to meet current and future production requirements. Through the Advanced Requirements Division, the Directorate provides the overall technical direction and coordination for development programs based on advanced concepts and technology.
    - Data Automation Division, Office of the Comptroller--establishes overall staff guidance and policy for all ADP activities and provides guidance, planning, and assistance relative to ADP activities. This office also monitors ADP utilization and supports new acquisitions related to Data Automation Requirements.

2. REQUIREMENTS ANALYSIS

2.1 Problem Formulation. The objective of this study was to provide an assessment of advanced computer science technology for processing and mass storage methods to provide more efficient operations and support of digital data bases required for existing and future production needs of the Defense Mapping Agency Topographic Center (DMATC) through the 1977-1983 time frame. The study provides for the analysis of data base requirements and data processing needs together with current pertinent research and development needs to formulate an advanced system architecture capable of meeting long-term data processing and storage requirements. Within these overall objectives the study was to include the following tasks:

- Establish as a baseline, the current processing and data storage levels
- Translate defined requirements into projected ADP and storage capabilities needed to support the DMATC mission
- Perform an analysis of data files
- Develop viable ADP system configurations for performance of the stated requirements
- Address the transferability of existing software for the selected configuration

2.1.1 Concentration of Study Areas. While the overall objective of the study did not change during the conduct of the study, certain factors emerged which influenced the manner in which the problem was approached and the selection of the factors considered most significant to long-term system development. Among these influencing factors were:

- The existence of a DMA-sponsored project to address requirements for an upgraded central processing capability
- A recognition that the data processing environment of the DMATC includes major capabilities in addition to the UNIVAC 1108 processors identified in the Statement of Work
- A contract revision encompassing consideration of "Back-End Data Base Management Technology" and an expanded analysis of data files with regard to on-line storage requirements

In addition to these considerations, it also became apparent that efforts would have to be concentrated in selected areas. In the course of the study numerous efforts leading to improved data processing and data handling were evidenced. Many of these efforts represent major undertakings by themselves, and a detailed assessment of each program was beyond the scope of this study effort. Those factors considered most significant from a long-term growth point of view were selected for concentrated study as having the strongest influence on the systematic growth of the DMATC data processing environment.

2.1.2 Study Approach. The initial phase of the study effort was the establishment of the current baseline. This was initially provided by a presentation from DMATC personnel who provided summaries of data systems and data processing activities reflecting DMATC requirements. These presentations also included data processing support provided by the DMATC for the DMAHC. These briefings were followed by more detailed discussions with data system program managers and personnel from the Department of Computer Services. A parallel avenue of approach examined areas of data processing technology, both through literature research and direct contact with government and commercial activities.

2.1.3 Organization of the Study. The remainder of this section of the report provides an assessment of the current baseline environment, and an examination of projected requirements for data processing and data storage capability. Sections 3 and 4 of the report examine selected areas of technology which will influence architectural options. Topics included are current DMA-sponsored areas of research and development, an assessment of back-end data base management technology, and an update of current mass storage technology. Section 5 presents a synthesis of requirements and technology factors contributing to an advanced systems architecture concept and illustrates the manner in which recommended system improvements can be integrated into a functionally responsive ADP support operation.

2.2 Analysis. This section provides an overview of the current DMATC data processing environment and a summary of projected requirements for both processing and data storage. In general, the approach describes the hardware environment, the functional data systems, and the data storage requirements. A rigid adherence to these topics has not been followed because of their high level of interaction. It is felt that a better understanding can be reached by addressing system entities. Finally, the implications of a DMA long range objectives and feasibility assessment program, the Pilot Digital Operations Plan, are examined.

2.2.1 Data Processing Environment. The purpose of this section is to provide an overview of the DMATC data processing environment from a computer hardware point of view. While the original Statement of Work primarily addressed the central processor (UNIVAC 1108) resources of the DMATC, it was subsequently recognized that other components of the processing environment should be recognized as essential parts of the total environment. This overview will therefore include the central processors, the Burroughs B-3500 processor, an IBM 7094 processor, and a summary of minicomputers currently in use at DMATC. Figure 1 provides a graphical summary of these components together with an illustration of external communications links now in use.

2.2.1.1 UNIVAC 1108 Central Processors. The primary data processing capability of the DMATC is provided by two UNIVAC 1108 processors operating under EXEC 8. The two processors are configured as independent systems to allow total isolation to accommodate security classification requirements. The UNIVAC systems are operated, maintained, and supported by the Department of Computer Services.

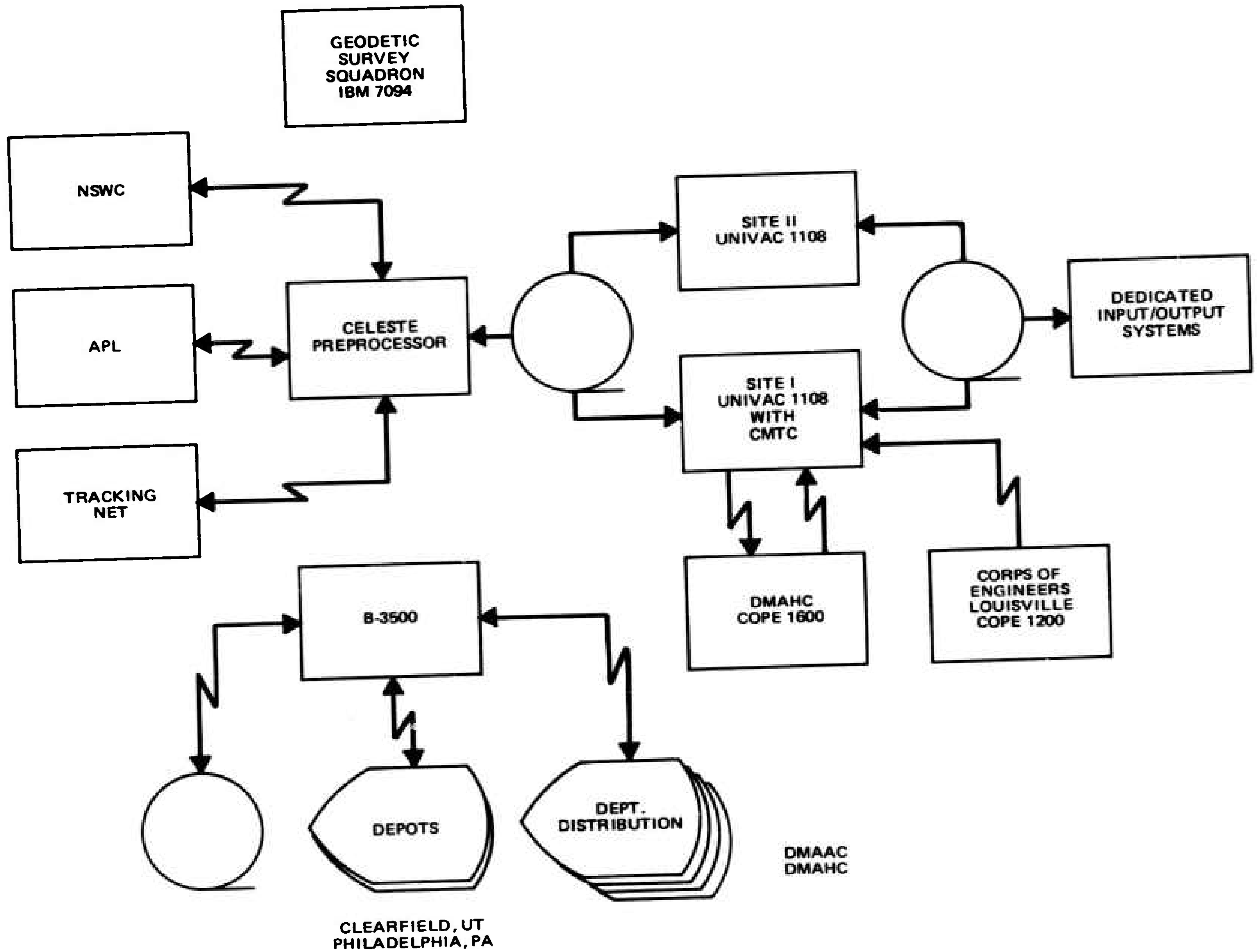


Figure 1. Data Processing Environment

2.2.1.1.1 Site I. The Site I UNIVAC 1108 is a 1 x 0 processor with 262K words of core memory. System peripherals include 4 drum storage devices (3 UNIVAC 432's, and 1 UNIVAC 1732); 10 tape drives; 10 model 8440 disk spindles; and other standard peripherals. The system also hosts a communications subsystem, consisting of a 32 port Communications Terminal Module Controller (CTMC). The CTMC provides a communications interface for:

- Five Uniscope 100 Communications Terminals
- An RM-1004 Remote Job Entry Communications Terminal
- A COPE 1200 Terminal located in Louisville, Ky. (Corps of Engineers)
- A COPE 1600 Terminal located at the DMAHC in Suitland, Md. This dual port terminal is supported by a 4800 baud link incoming to DMATC and a 9600 baud link for output to the DMAHC.

2.2.1.1.2 Site II. The Site II system also has 262K words of core memory. Peripherals include the same drum configuration as the Site I system; a Uniservo 16 tape subsystem with eight 9-track and four 7-track drives; and 6 model 8440 disk spindles. Additional peripherals include a high-speed printer. The Site II system has no external communication interfaces.

2.2.1.1.3 System Operation. The work load on the two UNIVAC systems is allocated on the basis of security classification with Site I providing primary processing for unclassified and collateral data and Site II providing primary support for SAA production. A concerted effort has been made to maintain a consistent minimal hardware configuration to allow flexible use of either system for backup or to balance system work loads. Similarly, standard software is maintained on the two systems to support flexibility in work allocation. Both FORTRAN and COBOL are supported on the systems with FORTRAN representing about 60 percent of all programs.

Both the Site I and Site II systems are operated on a nominal 5-day, 3-shift operation. All processing is done in a batch mode. Time block allocations are used for processing programs requiring extensive processor resources. Time block allocations are also based on security classification. Changes in the level of security are accompanied by system isolation and purge. A continuous attempt to obtain optimum operating efficiency through the monitoring and scheduling of time allocations for the two processors was noted. Extensive job accounting statistics are generated to continuously evaluate system performance and utilization.

Program related processing will be discussed later in this chapter.

2.2.1.2 Dedicated Input/Output Systems. An extensive use of minicomputers has existed at DMATC for some time. In most instances these computers function as dedicated control devices supporting a specific input/output system. A partial list of these devices and the associated minicomputer includes the following:

- Photogrammetric Devices
  - -- 7 UNAMACE (BR133)
  - -- 10 AS-11A (Bendix AP-2)
  - -- ARME (Varian 620f-100)
  - -- RPIE (PDP 11/45)
  - -- OLOP Off-Line OrthoPrinter (BR133)

- Digitizers
  - -- DPC--5 stations (PDP 11/20)
  - -- CALMA Disk System (Nova 1210)
  - -- CALMA Tape System (Nova 1210)
  - -- CEC--2 systems (Nova 1210)
  - -- DGR--10 stations (CDC 1700)

- Plotters
  - -- Xynetics (HP 2100A)
  - -- Concord II (Nova 1200)
  - -- Calcomp--4 systems
  - -- Automatic Type Placement--(SNAPS)--Concord I (PDP-8)

- CELESTE Preprocessor--Interdata 7/16

In addition to these items, new equipment acquisitions will add new minicomputers to this list of processors, such as the PDP 11/45 processor in the DIODES system, and the PDP 11/40 used in the DMA scanner-plotter system. None of the dedicated input/output systems are directly interfaced to the UNIVAC 1108 processors. Magnetic tape is the common interface media.

The minicomputers in use at DMATC illustrate an existing diverse environment. This diversity imposes requirements for support programmers well versed in a variety of programming languages, operating systems, and hardware characteristics. From an information management point of view, it also illustrates a need for data format standards for use in new system procurements. Use of such standards during the design process will preclude unnecessary processing work loads to reformat data.

2.2.1.3 Burroughs B-3500. The Department of Computer Services also operates a B-3500 computer system in support of DMATC operations. Although this system was not identified in the Statement of Work, the B-3500 supports essential mission functions and is an integral part of the total DMATC data processing environment.

The DMATC B-3500 was acquired from assets supporting the Air Force logistics support system. The B-3500 was first introduced in 1967 and is a medium-scale, third-generation processor. The system is business-oriented. The DMATC system has an extended core memory, and is supported by online disk storage, tape storage, and normal peripherals including line printer and word reader. The system also includes two telecommunications terminals to interface four remote query CRT display systems supporting the DMA Automated Distribution Management System (DADMS). Operating system software support for the B-3500 is provided by the Air Force Data System Design Center. This agency also provides and supports other standard base level software packages in areas such as financial management accounting and facility engineering.

The primary system supported by the B-3500 is the DADMS. The DADMS is operated by DMATC as a service to all DMA production activities in filling the product distribution requirements of the Department of Defense. The two primary operating shifts on the B-3500 are directed toward support of the DADMS. The configuration of the B-3500, including the remote query terminals, also reflects the orientation of the computer system toward support of the DADMS. Remote query capability exists between the B-3500 facility and distribution activities at DMATC, DMAAC, DMAHC, and the DMA map storage depots at Philadelphia, Pennsylvania, and Clearfield, Utah. This remote query capability and the use of leased telecommunications facilities for support of DADMS represent unique capabilities within the DMATC environment.

DADMS is a batch-oriented system which reached operational status in June 1976. The system consists of approximately 260 separate programs written in COBOL with some machine language code. The system is designed around the card-oriented military logistics system. Requisitions for DMA products are received from various military activities via card transceiver on AUTODIN communications lines. These requisitions are accumulated on tape for subsequent processing. Processing includes necessary subscriber and product verification and sorting for retransmission to the appropriate storage depot. The total system operates in the manner of a supply distribution system with appropriate provisions for inventory control, requisition suspense accounting, transaction history, etc.

DADMS activity is high and reflects the handling of some 120,000 requisitions, about 25,000 maintenance transactions, and some 70,000 automatic distribution addresses per month. Conversely, DADMS is relatively stable in size with an estimated growth of 5 percent per year.

Other systems supported in the B-3500 include comptroller-oriented financial management files including the DMIS/F and facility engineer files. These files are generally small and are projected to remain so in the future. While data in these financially oriented files is highly volatile, activity against these files is relatively low, i.e., less than once a day. A new file, the Advanced Personnel Data System-Civilian (APDS-C) is estimated to be operational by the end of 1976.

2.2.1.4 IBM 7040/7094. The DMA Geodetic Survey Squadron located at F. E. Warren AFB, Cheyenne, Wyoming, was recently transferred from the DMAAC to DMATC. This transfer consolidated DMA survey resources under the newly established Department of Geodesy and Surveys. Data processing resources of the Geodetic Survey Squadron are included in this study to provide a comprehensive overview of the total data processing environment of DMATC.

ADP capabilities of the GSS are functionally oriented to the editing, verification, and consolidation of field survey data collected by the field survey elements and detachments of the Squadron. The Squadron uses a variety of small processors, such as the Hewlett Packard HP-65 Programmable Calculator, to verify survey results on field location.

Final data verification and formatting is provided by an IBM 7040/7094 processor (Direct Coupled System) located at Cheyenne.

The processor configuration includes an IBM 7040/7094 (Mod I) Computer with 32K Core; 9 tape drives, 2 disk units with a capacity of 9 million 36-bit words; a Xynetics plotter; and other I/O peripherals.

In a typical reporting period, the system processed 387 jobs during approximately 75 hours of wall clock time during a one-week period. While the system is normally available on a 1-1/2 shift basis, it is available for contingency requirements on a full-time basis. The system is supported by a relatively small library of programs written in ANSI FORTRAN IV and Assembly language (MAP).

Subject to any significant change in survey requirements, a processing capability in the range of 75 to 100 hours (wall clock) per week appears adequate to meet projected needs of the GSS.

Of significance to the long-range projection, however, is the possible discontinuance of IBM maintenance and software support for the IBM 7094 class of processors in the 1979-1980 time period. This fact will require an assessment of alternative options to assure a sustained capability to meet the mission requirements of the GSS. This assessment should be initiated immediately to allow adequate assessment of unique processing requirements of the GSS geodetic applications work load--such as accuracy requirements (12 digits of significance) and the ability to solve large linear systems. The assessment should also include consideration of data management and storage requirements as they pertain to the GSS operational files.

2.2.2 Central System Utilization

2.2.2.1 Processing Activity. Detailed accounting statistics for the two UNIVAC 1108 systems were made available to the project team. Average utilization of the two systems during FY 76 consumed an average of 990 SUP hours per month. Typical level of activity is illustrated by the following average statistics for the Site I processor. The data reflects monthly averages from January through July 1976.

Cards In	3,381,000	
Cards Out	213,000	
Pages	417,000	
Tapes Used	6,194	(290/day)
CPU Time (Hours)	254	
SUP* Time (Hours)	558	
Jobs	6,961	(325/day)
CPU Time/Job (Minutes)	2.20	
SUP* Time/Job (Minutes)	5.10	
Jobs with Error Termination	1,507	(21.6%)

*SUP--Standard Unit of Processing. A unit devised to provide a consistent measure of processing service as viewed by the user program. Input to the calculation of SUP's is weighed such that SUP's will, as nearly as possible, determine elapsed time to perform a unit work in a serial environment on a unit processor with no overlap of I/O and CPU operations.

Utilization factors for the Site II system are similar and reflect system usage of approximately 80 percent of the Site I system. Significant factors include an average CPU utilization of approximately 200 hours per month, an average of 250 tapes per day, and approximately 220 jobs per day. The average CPU time per job is higher at about 2.6 minutes. The ratio of SUP time to CPU time is about 2.2 for both systems, and is generally representative of activity as illustrated in other accounting statistics examined. In general, it typifies the processing work load as a relatively high volume of short processing runs.

2.2.2.1.1 Disk Utilization. Under the current mode of system utilization, use of removable disk packs is discouraged by the Department of Computer Services. The standard configuration makes all disk spindles "fixed," i.e., system disks. There are no removable disk packs on the system except during specified "block" times when a system redefinition (SYSDEF) expressly produces a reconfiguration of the Site I system for a limited number of spindles. The normal use of all disk spindles as system disks in a fixed mode allows for and promotes a high degree of temporary file activity. Associated with this mode of operation, the operating system (EXEC-8) apparently encourages staging tapes to disk, presumably for update cases. The resulting typical sequence is frequently:

- Roll tape into disk (copy)
- Process file
- Roll data back to tape

In an attempt to obtain a clearer picture of loading relative to the level of system input/output (I/O) operations, a special collection of data was provided through the use of the UNIVAC Software Instrumentation Package (SIP). Similar systems data is being used by the Department of Computer Services to assist in work load allocation and the assignment of block times to achieve an optimum work load mix.

2.2.2.1.2 Evaluation of SIP Statistics. Approximately one week's worth of SIP statistics were compiled for each site. In order to provide a complete basis for analysis, "draw reports" reflecting memory usage and job contention for each day in which SIP monitoring occurred were also provided. All six SIP status codes were "ON" so the most complete processing profiles would be provided. The only variable then was the length of the period for statistics accumulation. This averaged a little over five hours on any given day while the draw reports on core usage usually covered one or more full shifts.

The following table reflects the periods during which valid statistics were taken by site. Note that there is no overlap between the two sites for any data accumulation.

Site	Date/Time Span					
I	7-26/4:22	7-28/2:03/5:30	7-29/4:04	7-30/4:26	8-03/7:46	
II	8-13/6:57	8-17/7:02	8-18/6:45	8-20/3:45	8-23/7:04	8-25/6:20
	8-26/2:04	8-30/5:47				

Particular attention was given to the SIP measures reflecting I/O activity, CPU usage categories, core usage, and job mix. These statistics were viewed relative to the profile histograms of the draw reports.

Analysis of this material showed that basically the utilization of both machines is highly efficient, especially in terms of the distribution of I/O activity across devices. Particularly, the amount of time that requests are queued for a channel are well distributed, although it is biased toward heavy tape activity. It was also noted that the following trends seem to prevail concerning the size of transfer blocks read (I) or written (O) on Site I devices (SIP measures in average word count):

Disks:	$I \approx O$	i.e., not much variation
Drums:	$I > O$	uniformly so
Tapes:	$I \ll O$	with few exceptions

No hypothesis could be formulated as to why this might be the case; no relationships could be established between this phenomenon and any other measure. On Site II, the relationship between input and output block sizes was much more uniform regardless of the device type.

It was also noted that the percentage of time the CPU is engaged in executive activity is directly related to the amount of time-sharing that is done. (Particularly evident in comparing Site I versus Site II since the latter supports no time-sharing terminals). Since time-sharing is heavily supported by executive level software, this can be interpreted as a more efficient use of the machine as such software is generally highly optimized. This implies a greater cost-effectiveness of resource usage when compared with standard batch compiler and debug methods.

Relative to overall loading, it was noted that both machines registered a low CPU idle time of less than 10 percent during the heavy use periods of prime shift. During second and third shift operations job loading histograms in the "draw reports" indicated that job turn around should always be less than 24 hours since slack or completely idle periods with no jobs in the machine were recorded for each day that statistics were tabulated.

2.2.2.2 System Utilization by Program Area. This section adresses the pattern and level of utilization of the UNIVAC 1108 systems. Data reflecting system utilization patterns was provided by the Department of Computer Services which made available monthly job accounting statistics for the two systems. The objective of this phase of the effort was to attempt to relate system processing activity with data file or product requirements. Use of job accounting statistics was only partially successful in this respect.

The primary method of sorting job accounting data is through use of DMATC organizational codes. Activity reflecting the number of runs, CPU usage, tape mountings, etc., is summarized on a daily and monthly basis. An initial examination of this data revealed that the use of organizational codes would not provide data useful to our purposes, since the data reflected the job

submitter's organization. Job accounting for all of FY 76 on this basis showed that departments used the UNIVAC 1108 resources as follows:

Computer Services	44.8%
Technical Services	6.6%
Geodesy	12.4%
Cartography	35.2%
Other	1.0%

A second type of job sort is based on the DMA program structure. This accounting method allows a sort to the three-digit level which identifies a specific product category such as 1:50,000 Scale Line Maps. Summary data at this level included SUP and CPU usage only. This summary was used to obtain a general relationship between data files or processes and computer work load. The relationships used are recognized as an over simplification since a given product may include processing involving multiple files or processes, but the dominant file relationship is considered valid. Utilization factors for the most significant users of the UNIVAC 1108's is summarized in Table 1 and graphically portrayed in Figure 2. The dominant factor apparent from the summary is that three program line items--the 1:50,000 Line Map, the 1:50,000 Topographic Data Base, and the Digital Topographic Elevation Data--consume 45 percent of the total resources. The dominance of these three interrelated product areas is consistent with the Center's mission. A secondary observation is the amount of resources directed toward resource management which approaches 40 percent.

A general observation made during the examination of the job accounting statistics relates to the ratio of jobs with error termination to the total number of jobs processed. Use of this factor can be an effective tool in exercising quality control provided program runs are identified as to status, i.e., production, pre-production test, or developmental (programming and debug). Use of standard program designators to identify both the pertinent data system and program status can be an effective means of identifying high usage programs which should be maintained on the system library, programs having high error termination, or programs which should be examined for efficiency. Identification of production programs having a high error termination rate can be used by quality control personnel to isolate problem areas and initiate a trace for the source of the error. On production-tested programs the most frequent source of such problems is in the data preparation phase. In a batch mode operation, such terminations have their most adverse effect on the production pipeline since the job submission must be returned to the user, corrected, and resubmitted.

Implementation of standard computer program designators together with an identification of their production status could provide a useful aid to quality control functions examining the production process.

2.2.3 <u>Data Systems.</u> One of the baseline considerations for this study included the major data systems in operation at DMATC. Specific systems identified for consideration include the following:

Table 1. UNIVAC 1108 CPU Utilization
(FY 76)

Line Item	Includes	CPU Hours	Percent
3AA 1:50,000 Line Maps	SACARTS	380	7.0
3AC 1:50,000 Topo. Data Bases	DARIC	1,037	19.2
3EA Dig. Topo. Elevation Data	DTED	1,004	18.6
3GE DOD Library of Maps	TDLS	246	4.6
3GG Analysis and Evaluation	PMS	188	3.5
3GJ Procurement of Maps and Data	ACRES	240	4.4
4FD Geodetic and Gravity Studies		225	4.2
4FF Satellite Geophysics Program	CELESTE	296	5.5
7BA Mission ADP Support		185	3.4
7BC Product/Techniques Development		384	7.1
7BD Program Management Info. Sys.	DMIS/P ARAPS	299	9.5
7BE Other Mission Support		515	5.5
Others		396	7.5

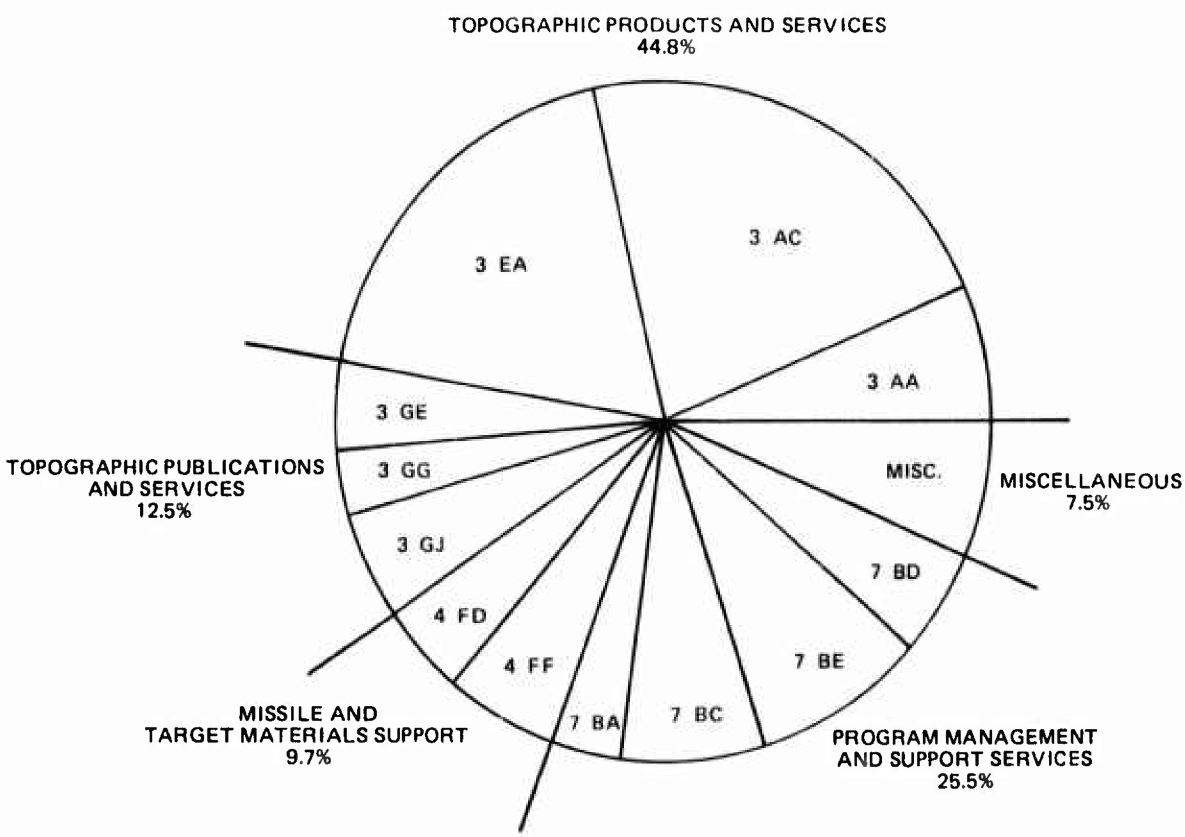


Figure 2. Utilization of U-1108 Resources by Program Area FY 76

- Topographic Data Library System (TDLS)
- Semi-Automated Cartographic System (SACARTS)
- Data Reduction Interface and Control System (DARIC)
- DMA Management Information System (DMIS/P)
- Digital Topographic Information Bank (DTIB)
- Product Management System (PMS)/Area Requirements and Production Status (ARAPS)
- Automatic Map Research Report (AMRR)
- CELESTE

This section contains a brief summary of these systems together with statistical data which reflect a summary of file size and activity.

2.2.3.1 Topographic Data Library System (TDLS). The TDLS is an automated library which describes holdings of maps and related textual data. The TDLS is operated as a consolidated library for the Department of Defense and serves as the single reference for storage and retrieval of published maps and related data. The TDLS supports production requirements of all the DMA activities as well as other agencies in the Federal Government.

The TDLS constitutes a merged library of three formerly manually maintained data files which included maps, books, and intelligence information. At present the system provides a summary directory to 1,700,000 maps and associated textual documents; 40,000 books; and 40,000 other documents. The loading of the TDLS data base was completed during the fall of 1976. The TDLS produces a variety of products and reports including a Master File, Master File Deletion Report, Statistical Summary, Document Summary Cards, etc. The file includes classified data since indexing and references to holdings are provided the same protection as the original source material.

Programs supporting the TDLS on the UNIVAC 1108 are written in Level 14 COBOL. The file contains some 72 million characters and is stored on approximately 50 magnetic tapes including backup. To reduce computer time involving file maintenance activities the TDLS has also been placed on disk.

Following initial implementation and system loading, update activity is projected at about 5,000 entries per month with an expected 40 to 50 queries per month. Although the system has been operated on a daily basis during the initial loading of the data base, there is little current experience on which accurate computer usage factors can be established for the operational system. This will be largely dependent on the extent to which DMATC cartographers utilize the data base as a resource in assessing the availability of map source material for new production requirements.

2.2.3.2 Semi-Automatic Cartographic System (SACARTS). The SACARTS consists of the combination of both hardware and computer software which support DMATC cartographers in the production of reproduction quality materials. The system has been in development for an extended period of time and has been introduced into the production process over the past two years. Used on limited production basis during this transitional period, the SACARTS presently represents a mixture of automated and manual functions depending on specific functional capabilities and an optimal selection of techniques best suited to supporting map production.

The production work load which will be supported by SACARTS has been programmed from its initial rate of 50 charts per year to a production rate of 500 charts per year. Because of this heavy production dependence on the system, SACARTS was isolated as a subject of special interest and was investigated in greater depth to help identify system choke points and to provide recommendations pertinent to long-range system architectural plans. The summary of this overview is presented separately in Appendix A.

Computer processing supporting the SACARTS is provided in a batch mode on the Site II UNIVAC 1108. Use of a large scale computer is central to the original SACARTS design concept. Data storage related to SACARTS files is provided by standard magnetic tape. Approximately 200 archival tapes are maintained at the present time plus an additional 80 tapes containing planimetric data held as part of the planimetric subsystem of the Digital Topographic Information Bank (DTIB). The data bank of planimetric data is expected to grow at a rate of 500 tapes per year. Present plans provide for maintaining these tapes in the SACARTS related Graphic Improvement Software Transformation System (GIST) format.

Development of a DMA feature classification and format standard for planimetric data will require the development of transformation software to allow reformatting GIST data to the DMA standard.

2.2.3.3 Data Reduction Interface and Control System (DARIC). DARIC can be most accurately described as a system or process which supports the extraction of topographic information from photographic source materials. All map compilations for which basic terrain information is derived from photography are supported by the DARIC system.

DARIC provides the means for managing, storing, and transferring geodetic triangulation data from one function to another within the DMATC environment. As in most geodetic operations, the system operates under stringent numeric control. In a simplified sense, DARIC is concerned with the translation of positional information in a photographic image to its true earth position. Elements entering into the process may include factors such as camera station, camera calibration data, the identification and measurement of known ground control points within the image, and other factors. From this data orientation parameters can be derived for use in a stereo-plotting instrument such as the UNAMACE or the AS-11A. The process also supports analytical triangulation for the extension of ground control information.

The establishment of photographic orientation parameters is a highly tailored process dependent on type, quality, and quantity of input information available. The DARIC system supports this process through a library of programs which may be selectively used to take best advantage of available data. The complexity of the process involving source selection, image transfer, photographic measurement, and supportive computer processing involves extensive lead time.

The DARIC system provides triangulation data to support 7 UNAMACE and 10 AS-11A analytical stereo-plotting instruments. The output of the UNAMACE consists of a uniform grid of elevation points. Following a coordinate transformation process on the UNIVAC 1108, a master tape is created which constitutes a permanent record of the terrain data extraction process and is the source record for subsequent data formatting processes.

Since the DMATC program is primarily directed toward the production of line maps, terrain data is formatted in contour form. The Contour III program incorporating advanced terrain modeling features is currently in use. (A high run-termination rate has been expereinced in conjunction with the contouring function, but is attributable to errors in preparation of the run deck rather than the program itself).

Although the UNAMACE master tapes hold the potential for direct production of digital terrain elevation data in matrix form, they are not presently exploited in this manner. Editing functions performed during the contouring process and the entry of geomorphological corrections made during the UNACOMP process are required before an accurate contour manuscript can be prepared.

DARIC related programs are run on the Site II computer. While the process itself is a tailored operation making best use of available source input, the resulting UNAMACE master files reflect a stable (non-volatile) data base asset. This data base consists of approximately 2,500 magnetic tapes and has a projected growth rate of approximately 18 percent per year.

2.2.3.4 DMA Program Management Information System (DMIS/P). The DMIS/P is a major subset of the overall DMA Management Information System (DMIS) and should be addressed within the overall context of the DMA information management concept.

2.2.3.4.1 DMIS. The DMIS is comprised of a set of interrelated functional management information systems. Each of these functional systems is structured around its own distinctive requirements and data sources, but must mutually support the total information system through the exchange of basic data and informaton displays.

The objective of the DMIS concept is to integrate these functional management information systems into a cohesive structure that can facilitate communications and support the decision-making process at all levels of DMA management. One of the more specific system objectives of DMIS is to provide an information system that supports and integrates the planning, programming, budgeting, and execution review process within the DMA.

The DMIS establishes the policy under which information is made available at the various management and control levels within the DMA. Consequently, DMIS places stress on information management rather than on the automated data processing systems required to generate and support the information system. Information within the DMIS structure is gathered and processed at various functional levels and may be supported through a variety of programs and data handling processes both manual and automated.

The DMIS is structured to provide information in five categories, all of which contribute to an overall assessment of resource management. The five subsets are:

- Program Management--DMIS/P
- Financial Management--DMIS/F
- Equipment Procurement Management--DMIS/E
- Support Management--DMIS/S
- R&D Management--DMIS/R

The most complex of these systems is the DMIS/P which not only collects information relative to production center production programs, but also provides an interface at the headquarters level to four related data files:

- DMA Resource Objective Plan (DROP)
- ACRES
- Area Requirements and Production Status (ARAPS)
- Productivity Measurement System (PMS)

In addition to maintenance of the Center's DMIS/P, the DMATC also supports the maintenance and exploitation of the ACRES, ARAPS, and PMS for Headquarters, DMA.

2.2.3.4.2 DMIS/P. The DMA Program Management Information System (DMIS/P) is designed to portray manpower resources against mission and mission support requirements. The system is oriented to support the program formulation, resource allocation, and program execution reviews at successive levels of DMA management. The required information is accumulated at component levels via programs and available resources at the three production centers. Consequently, the DMIS/P at the DMATC is based on use of the APPMIS and is supported on the UNIVAC 1108 processor; the DMAAC DMIS/P uses the PROMACS system on the Burroughs B-3500 processor; while the DMAHC system is based on the FAMIS system supported by a WANG processor.

A significant feature of each of these systems as well as the entire DMIS concept is the use of a standard DMA program structure which established a common product (or service) oriented information structure for the allocation and reporting of resources. The use of this common product oriented structure is considered a significant basis of commonality on which future growth and expansion of the DMIS can be structured.

2.2.3.4.3 The DMATC DMIS/P. The DMIS/P consists of approximately fifteen files maintaining program schedules, programmed resources, actual resources used, product schedules, and similar data. The individual files are maintained on 20 magnetic tapes and contain a total of 615 million characters of data. The files include programs for transfer of map production status to the ARAPS. The system also allows direct entry of data from related management information systems maintained by the Center's production departments. Data from the Field Offices is collected by the Louisville facility and transmitted via a Corps of Engineers COPE terminal to the DMATC UNIVAC 1108 facility. The DMIS/P involves use of approximately 40 programs supporting the Center level MIS and an equal number supporting the complementary department level systems.

The DMIS/P is projected to remain stable in size. File content is highly volatile reflecting weekly transactions involving approximately 3,500 time cards charged against 2,500 projects. While the system reflects a weekly update, multiple operations and validation of input data (time cards) require processing on a daily basis. The system also supports a variety of management reports with summary staff reports generated on a monthly basis. Data for entry to the ARAPS is prepared on a quarterly basis, and reflects the only direct tie with other resource management files.

The DMIS/P is totally supported on the UNIVAC 1108 in batch mode and involves both unclassified and classified data.

In addition to the DMIS/P itself, individual departments support and operate departmental Management Information Systems reflecting resource management and production scheduling. For example the "Carto Line System" (CLS) is operated by the Department Field Services. Inputs to the CLS are prepared in the Field Offices and mailed to the Louisville Field Office for consolidation. The consolidated input is transmitted to the DMATC UNIVAC 1108 facility via services provided by the Louisville District Corps of Engineers COPE 1600 terminal. Processed data is returned to Louisville by the same route and distributed by mail to the other Field Offices. The system is updated biweekly and reflects an input activity of about 750 cards.

2.2.3.5 Digital Topographic Information Bank (DTIB). The DTIB is a digital data bank which is being developed by the DMA to serve a variety of production requirements and users. Development of the DTIB is being coordinated by the DMA with participation of the three production centers.

Digital Topographic Data is being addressed in two subcomponents, Digital Terrain Elevation Data (DTED) and Digital Planimetric Data. The DTED is in an advanced development stage with the DMAAC serving as Executive Agent for the system development. Planning for the Digital Planimetric Data Base is proceeding with the development of a DMA standard for the definition of a feature classification scheme and data format standard. Coordination and adaptation of a DMA standard is anticipated early in calendar year 1977.

The development of the DTED as a standard DMA data file is based on use of the "DMA Standard For Digital Terrain Elevation Data File." This standard establishes a uniform structure for the content and format of digital terrain elevation data. The file structure is organized into 1^{o} x 1^{o} geographic sequential areas using a southwest corner origin. Elevation points are arranged in a south to north profile, sequenced from west to east.

The horizontal plane spacing of the elevation matrix is in whole second intervals for intervals of 1 second or above and in multiples of 0.01 second for intervals less than 1 second. Data storage standards are based on the use of 9-track 1600 bpi magnetic tape. Data for storage will consist of ten 1^{o} x 1^{o} blocks per tape. The DTED software support system is being developed at DMAAC and uses the UNIVAC DMS-1100 data management system for implementation of an automated directory. The full system is targeted for completion in early 1977 and will include:

- An automated file directory
- Report generation capability
- A file management system (FMS) to support data extraction, file merging, and file extension (beyond 1^{o} x 1^{o} blocks) features
- Automated transaction log

The DTED system was installed at DMATC following system test.

Current actions at DMATC are being directed toward the conversion of existing matrix elevation data to the DMA standard. Current elevation data held at the DMATC exists in three formats:

- DMA standard with a grid interval of 3 seconds
- Planar Tapes. These tapes contain data from 1:250,000 map source. Data is recorded at 0.01" in table coordinates in a structure similar to the DMA standard. Each 800 bpi tape reflects data from one map sheet.
- UTM-Planar. These 800 bpi tapes record data from 1:50,000 and 1:25,000 topographic maps in a UTM coordinate reference. The ground interval for the 1:50,000 scale map sheets is 12.5 meters.

Holdings currently consist of 200 tapes in DMA standard format and approximately 1,500 tapes in Planar and UTM-Planar format. In addition, approximately 6,000 tapes containing Planar working data are being screened for verification, compaction, or elimination. Despite a significant growth projection for data in the DMA standard format, the combined effect of file growth and tape management is expected to reduce the number of tapes that will be held in 1979.

Computer processing associated with the DTED is projected to require approximately 100 hours of UNIVAC 1108 CPU time per month. Approximately 90 percent of this work load is associated with the generation and formatting of data in matrix format while an estimated 10 percent of the work load is associated with maintenance of the automated DTED data bank.

Data content in the DTED file is expected to remain highly stable within a completed tape holding. Activity in the data base will be initially characterized by file building activity with a gradual transition to file exploitation as the data base plays a more active role as an input source supporting the production of DMATC end products.

2.2.3.6 Product Maintenance System (PMS) and Area Requirements and Production System (ARAPS). The PMS and ARAPS are independent systems which both relate to the status of map production. The PMS data files are used to maintain currency and accuracy evaluations, while the ARAPS essentially provides a record of the product requirements cycle.

The Area Requirements and Production System is a system used to record and consolidate map production requirements submitted by the operating commands of the Department of Defense and other appropriate government agencies. Map requirements are submitted to the DMA based on operations plans maintained by the Unified and Specified Commands. When consolidated, the total requirements are evaluated against the ARAPS production status file which records production in progress and the available status file which reflects completed production. Input from the PMS is also interfaced with the available status file to validate product currency and accuracy.

The ARAPS cycle closely follows the command's submission of requirements on an annual basis. System activity peaks around October with maintenance and updates entered on a monthly cycle. Priority requirements for Air Target Materials are processed on an as-required basis.

The ARAPS data files consist of 40 million characters and are maintained on 4 magnetic tapes. The file represents approximately 74,000 "available" items; an item requirement of 71,000; and a combined "requirement" and "available" file of approximately 38,000. The PMS file contains some 12 million characters maintained on 2 magnetic tapes. Programs for both systems are written in COBOL.

The ARAPS has been recently augmented with a graphics display module to facilitate the generation of composite graphics reflecting requirements; information on availability, adequacy, and currency of required items; information

on items programmed for production; and scheduled completion dates for items in production.

The ARAPS was initially developed for the DIA for the IBM 360 processor and subsequently converted to the GE 635 processor. The lack of full and complete documentation of the original system programs has been a severe constraint on full exploitation of the system by the DMATC.

2.2.3.7 Automated Map Research Report (AMRR). The Automated Map Research Report is a new file which is being considered for use by the Department of Technical Services. The file would provide an automated record of information currently contained in a manually maintained folder record. The folder is maintained on an area basis. These existing Map Research Reports (MRR) contain information useful during the production planning process to facilitate the identification and selection of the best means of supporting new production tasking. Examples of the contents of an MRR include:

- Availability of map coverage
- Accuracy of available map coverage
- Identification of best map sources for recompilation
- History of previous product development
- Area geography/geology
- Demography
- Construction, etc.

The MRR is analyst oriented and includes extensive use of graphics for portrayal of geologic formation, population growth patterns, and rainfall projections.

Planning considerations for an automated MRR have included the recognition that much of the data pertinent to the AMRR exists within other files and directories maintained at the Center. Informal estimates provided indicated that as much as 60-70 percent of the information required is already contained in files such as the:

- TDLS
- ARAPS
- PMS
- ACRES
- Geographic Place Names/Gazeteers

Additional information also exists in the Geodetic Library, but is not currently automated. A major portion of the report content not currently existing in digital form is expected to consist of geologic interpretation data.

Current efforts are being directed toward an assessment of the AMRR concept to determine the relative merits of proceeding with an automation concept.

2.2.3.8 CELESTE. CELESTE is a computer processing system used to determine precise orbits for earth-orbiting satellites. The system provides a means of maintaining and updating ephemeral information based on tracking data collected from approximately 40 Geoceiver stations distributed worldwide. The system supports orbit determinations for the TRANSIT navigation satellite and in the future will also provide similar support for the Beacon and GEOS-3 satellites.

The DMATC support of the navigation satellite programs consists of three separate functions:

- Consolidation, editing, and formatting of tracking data files
- Computation of the precise ephemeris (CELESTE)
- Computation of geodetic positions

Input tracking data for use in the CELESTE computations was previously provided by the John Hopkins Applied Physics Laboratory through a terminal at DMATC. Functional responsibility for consolidating and editing the raw tracking data has been largely transferred to the DMATC and is being processed on an Interdata 7/16 minicomputer supported by disk and tape storage, card and paper tape input devices, and two line printers. The processing system is also supported by a telecommunications interface for direct interface with the APL and the Naval Surface Weapons Center at Dahlgren, Virginia. Data processing support related to the editing of tracking data has been incrementally moved from APL to DMATC. The remote interface to the APL will continue as this transition is completed to accommodate tracking data not received directly at DMATC. The direct entry of external source data to the CELESTE Interdata processor is unique in the DMATC environment.

The CELESTE system requires daily processing support from the UNIVAC 1108. This support includes daily short runs related to formatting of tracking data, a more extensive run daily for the generation of the precise ephemeris, and runs on an as-required basis for the computation of geodetic positions. Since CELESTE tracking input is essentially a continuous process, a continuous processing cycle is essential to avoid backlogs. While system response time requirements (7 to 10 days) do not impose stringent processing requirements, the continuous flow of input data essentially dictates a processing cycle consistent with the input data flow, i.e., less than 24 hours.

Expansion of CELESTE processing to accommodate seven satellites in early 1977 will place an increased work load on the UNIVAC 1108 processor. The increased level of activity is expected to triple UNIVAC 1108 processing requirements during peak seasons. Since the work load has significant seasonal variations,

a continuous assessment of work load projects will be required to minimize the central processor work load scheduling.

Data storage for the CELESTE system is provided by magnetic tape. Tape holdings currently include approximately 300 reels with a projected growth of archival tapes of approximately 60 reels per year.

This growth rate is considered nominal and will require a continuous tape management effort based on well defined standards for data retention. Establishment of an archival directory has also been recognized as a necessity for efficient storage and retrieval of ephemeral data to service user requirements.

2.2.3.9 Data System Summary. The data systems previously described represent the most active data systems in use at the DMATC. In addition to those described, the ACRES system together with a number of free-standing but related files were of pertinent interest from an activity point of view. ACRES is not further described for security reasons.

In addition to these files, a variety of other individual files and directories were identified. Some 50 files were identified, most of which represent relatively small data volumes. Approximately 10 of these files are supported on the Burroughs B-3500 processor.

A summary of file or data system characteristics was compiled. The summary reflects:

- Security
- Storage Media and Volume
- Projected Growth
- Level of Activity
- Response Time
- Volatility

Tables 2 and 3 illustrate these file characteristics for the primary systems described as well as selected additional files. While most of these features are self-explanatory, several deserve additional explanation. Level of Activity is intended to isolate files having high frequency processing activity. Response Time was tabulated to obtain a general sense of priority. Volatility is a general assessment of the stability in a given set of data, i.e., slowly growing directories will reflect low volatility while a file such as DMIS/P will reflect a high level of change on a weekly basis. These last three characteristics are of particular interest when evaluating files as candidates for online storage.

Table 2. Data Base Characteristics

File	Security	Storage Media		Projected Growth	Level of Activity	Response Time	Volatility
		Tape	Disk				
TDLS	Collateral	50	(4)	0	Daily	24 Hr.	Low
SACARTS	SAA	200	(2)	500 tapes/year	High/Daily	24 Hr.	Low
UNAMACE	SAA	2,550		18%/year	High/Daily	24 Hr.	Low
DMIS/P	Unclassified	20		Negligible	5/day	>24 Hr.	High
DTIB	Collateral						
DTED		1,100		200/year		24 Hr.	Low
Unverified							
Planar		6,000		Consolidation		24 Hr.	Low
Planimetry		80		500/year		24 Hr.	Low
PMS	Collateral	2		Negligible	1/month	24 Hr.	Moderate
ARAPS	Collateral	4		Negligible	Daily	24 Hr.	Moderate
AMRR	Collateral			(In Planning Stage)			Low
CELESTE	Collateral	300		60 tapes/year	High/Daily Seasonal	<24 Hr.	High
ACRES and related files	SAA	16		Total 8 tapes/yr	High/Daily	2-24 Hr.	High

Table 3. Additional Data Files and Characteristics

File	Security	Storage Media		Projected Growth	Level of Activity	Response Time	Volatility
		Tape	Disk				
Repromat DB	Collateral	2		?	1/month	24 Hr	Low
Gazetteer DB	Unclassified	175		?	1/month	24 Hr	Low
Interim Mag. Tape Library	Collateral SAA	1 1		3%/year	1/week	24 Hr	Low
U-1108 Accounting	Collateral	44		Stable	1/day	24 Hr	High
Analytical Extensions	SAA	151		?	3/week	24 Hr	Low
Country Files	Collateral	225		?	?	24 Hr	Low
Total		559					

2.2.4 Current Data Storage.

2.2.4.1 Magnetic Tapes. The primary method of digital data storage is provided through the use of standard magnetic tapes. The current inventory of tapes is estimated at approximately 25,000 reels, and includes the following general groupings:

- Data Files--including TDLS, DMIS/P, PMS, ARAPS, ACRES, and CELESTE (392 reels)
- Other Directories and MIS (631 reels)
- Source/Product Data Bases--including DARIC, SACARTS, and DTIB (9,930 reels)
- Others DC/S Tape Library, DMAHC tapes, etc. (4,000 reels)
- Scratch and Work Files (10,000 reels)

An active program to evaluate and validate magnetic tape holdings has been initiated under the guidance of the Data Base Administrator. The objective of the program has been to assure adequate protection and security for tapes which can support future production needs (source data base) as well as to release tapes which contain data of marginal utility or which reflect a temporary work status. Under the DTIB, Digital Terrain Elevation Data, is also being consolidated following a verification process with a view toward minimizing tape holdings and providing a more manageable means of tape storage and retrieval of this data. This effort will allow systematic retrieval of data for future production and recognizes the high monetary value of the data maintained in this consolidated form.

2.2.4.2 Disk Storage. As a general rule, disk storage associated with the UNIVAC 1108 system is used in a fixed system definition mode. Although a limited number of files are maintained on removable disk (principally the TDLS), this practice is considered an exception and is not encouraged. The utilization of disk spindles on the two UNIVAC 1108 systems is shown below and reflects the allocation of storage capacity between catalogue files (program files), temporary working storage (scratch files), and removable disk files.

Allocation of Disk Storage Capacity

(Millions of Words)

Site	Catalog Files	Scratch Files	Files on Removable Disk	Total System Capacity
I	98.5	40.7	68.3	200.0
II-SAA	51.0	28.5	0	120.0
II-Classified, NON-SAA	22.8	36.8	83.9	120.0

The above disk storage together with a limited amount of drum storage reflects the only online storage capacity of the UNIVAC 1108 systems at the DMATC.

2.2.5 Projected Data Processing Requirements. Projections of data processing requirements through the 1980 time frame were provided to the study team. These projections consisted of two elements: requirements based on approved production programs (validated requirements) and requirements that reflected anticipated needs (unvalidated requirements). The analysis was based on data processing requirements associated with specific product or service line elements of the DMA Management Information System Program Structure. This method of documentation allows a correlation between line item production requirements as well as security requirements associated with the production process. The projections included UNIVAC 1108 processing requirements for both the DMATC and DMAHC.

Estimates provided by DMATC are recognized as working estimates current as of January 14, 1977. As with all planning documents, the requirements presented will be subject to revision as production requirements are revised and as planned system augmentations in both the central processing systems and in other Center components such as the dedicated input/output processing systems occur. The format of this requirements document, based on DMA program structure line items, should prove a valuable mechanism for a continual assessment of data processing capabilities in a dynamic environment.

2.2.5.1 Overall Processing Requirement. The overall processing requirement projected by the DMATC is graphically shown in Figure 3. The figure shows total central processing requirements on an annual basis expressed in UNIVAC 1108 CPU hours. The baseline figure of 5,395 hours reflects actual CPU utilization for the UNIVAC 1108. This projection is based on the most recent assessment of work load based on current plans. The data is recognized as highly dynamic and is subject to continuous revision as production requirements change.

Projections for 1977 through 1980 reflect an overall processing requirement approximately twice that of the 1976 level based on validated requirements. Potential requirements, reflecting identified but as yet unvalidated requirements, show a further increase approaching a level three times FY 76 requirements during the 1978-1980 time frame.

The figure also shows the estimated capacity of the existing computer systems. These estimates, shown both in terms of a 3-shift, 5-day week operation and maximum capacity, are based on actual system utilization characteristics. The estimates account for lost time due to preventive and unscheduled maintenance and due to security purges. The estimate is then based on the ratio of CPU time to available wall clock time. Since this is based on experience data obtained in the DMATC production environment, it correctly portrays system capacity based on the actual mode of operation. While this does not exclude improvements which may accrue in specific programs or processing techniques, it does provide a realistic basis for future planning.

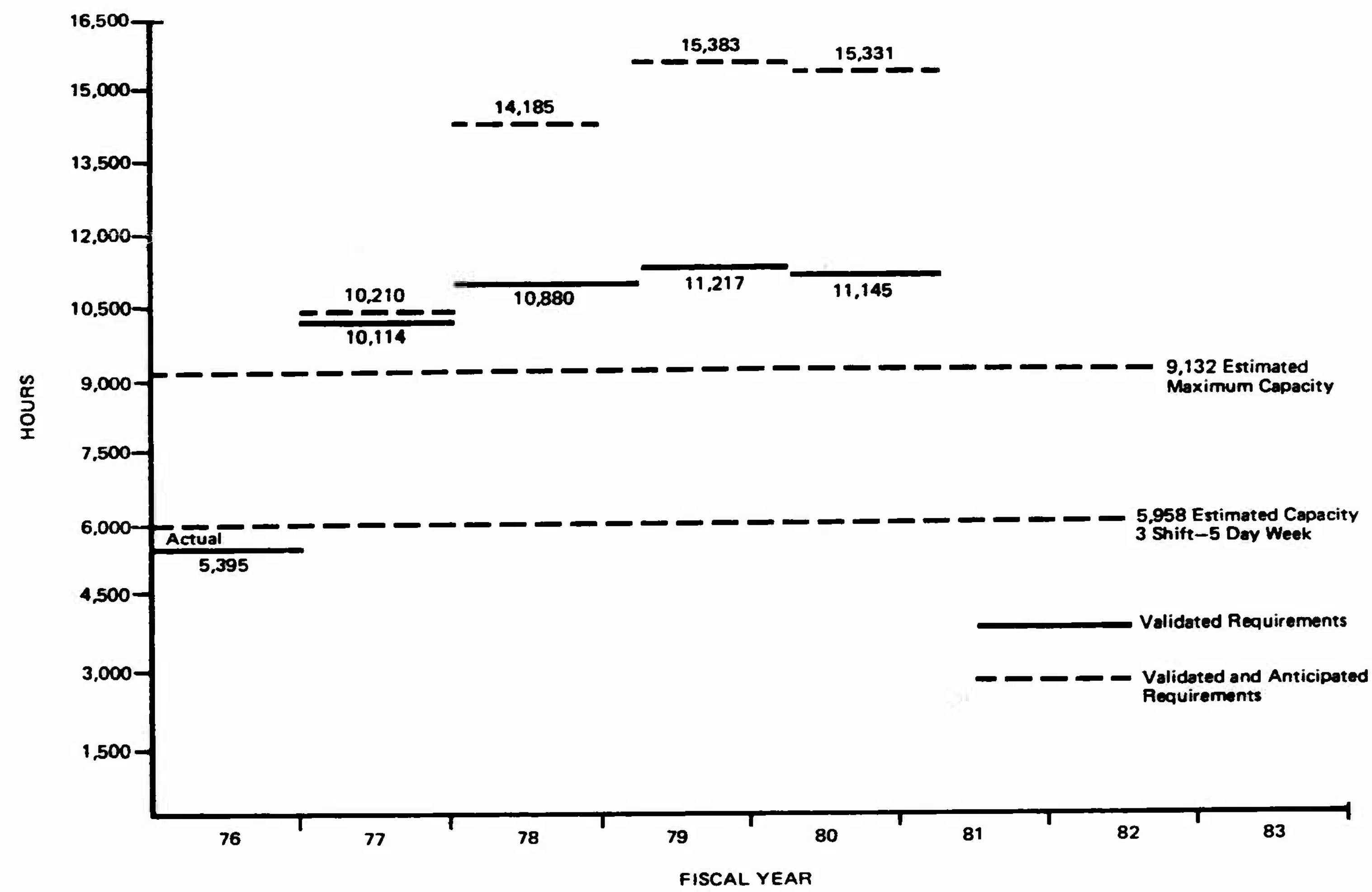


Figure 3. Total Projected Requirement (Hours of U-1108 CPU Time)

The most salient implications of the projection are the rapid increase in requirements in FY 77 and the significant overall growth which may be required by 1978. These factors directly affect system architectural options both in terms of time phasing, system capacity, and trade-off options.

2.2.5.2 Collateral Processing Requirement. Because of the dominant constraint that security classification has on the DMATC data processing environment, a further examination of the requirement was made from a security point of view. Figure 4 portrays the total data processing requirement based on the collateral work load. For this purpose, the capacity estimate is based on system utilization experience for the Site I system, which again provides a baseline calculated on the actual production environment.

The data processing requirement for collateral data reflects an increase in FY 77 which approaches the Site I system maximum capacity. Validated requirements are projected to slightly exceed maximum capacity beyond 1977, leaving no capacity for as yet unvalidated requirements. These unvalidated requirements reflect a potential total system work load approximately two times the estimated system capacity.

The major implications of the collateral work load are the:

- 40 percent increase in work load during FY 77
- Site I system at saturation starting in FY 77
- No capacity to accommodate anticipated growth resulting from unvalidated requirements beyond FY 77.

2.2.5.3 SAA Processing Requirement. A similar breakdown reflecting SAA processing requirements is shown in Figure 5. The rapid growth of SAA processing requirements is most apparent in this graphic with the most significant growth (148 percent) occurring during FY 77. The projection shows requirements exceeding system capacity occurring in FY 77 with more moderate increases occurring through 1980. Virtually all of the SAA processing requirement is associated with validated requirements.

2.2.5.4 Implications of Projected Processing Work Load. From a system architecture view the significant factors of the DMATC data processing work load study are:

- An immediate need for increased CPU capacity for both SAA and collateral production.
- FY 77 time constraints preclude an initial offload to distributed processing systems.
- The rapid growth of processor work load will increase tape handling by approximately 100 percent during FY 77.

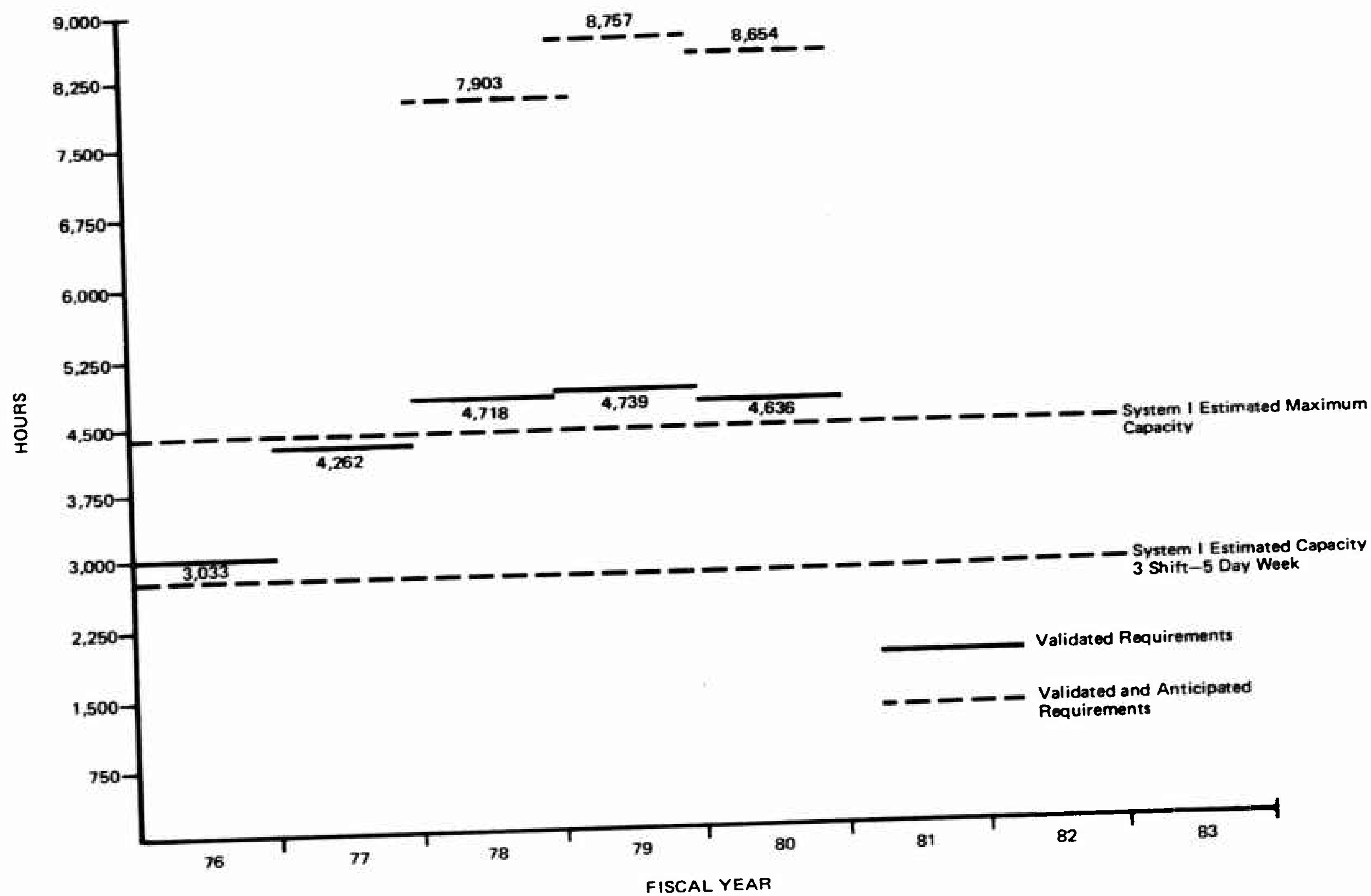


Figure 4. Projected Requirement—Collateral Work Load (Hours of U-1108 CPU Time)

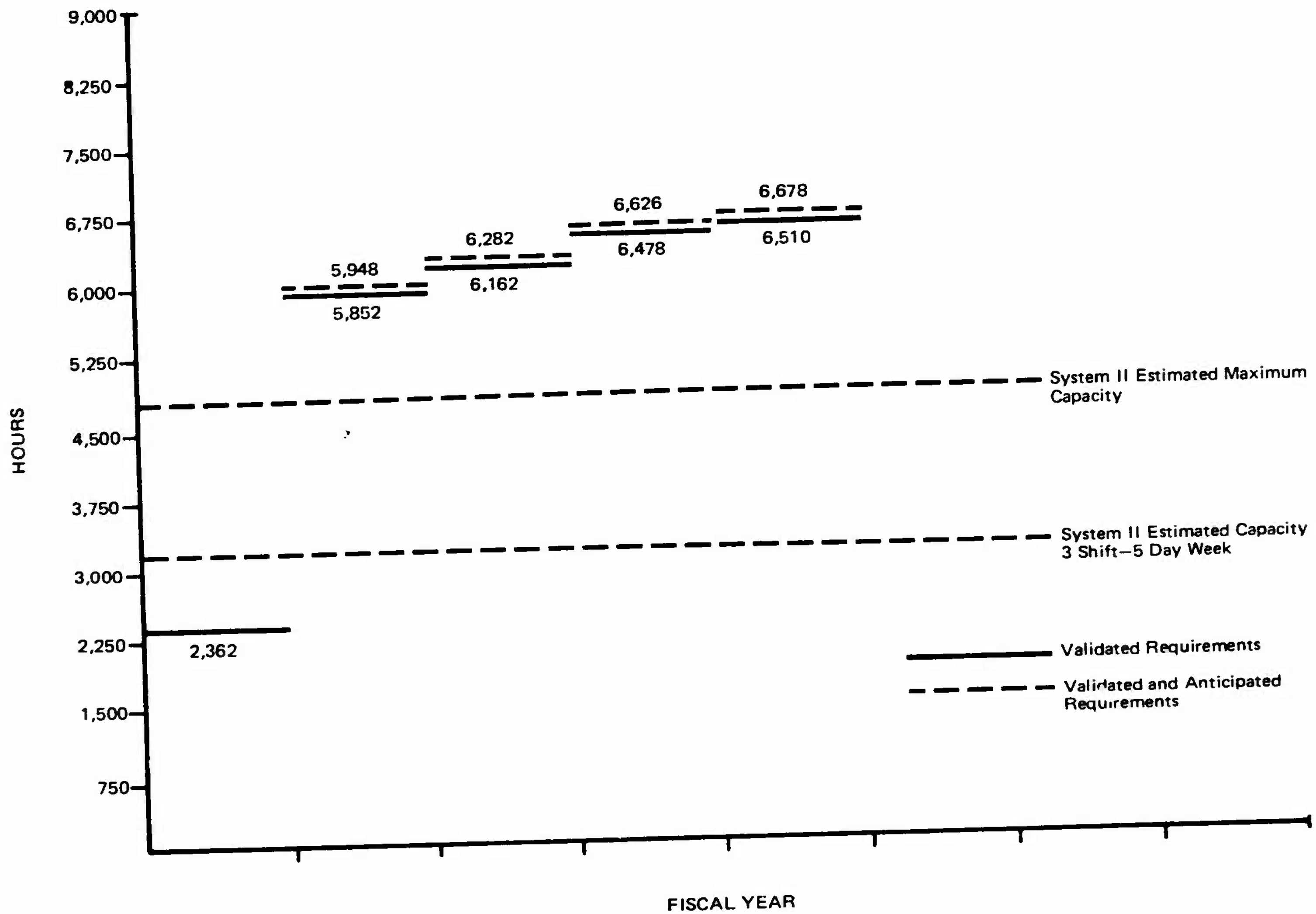


Figure 5. Projected Requirement—SAA Work Load (Hours of U-1108 CPU Time)

- System users can expect increased turnaround times as greater portions of the work load encompass second and third shift (or weekend) processing.

- Primary shift operations, supporting DMAHC and DMATC "open shop" users, will be severely impacted by the rapidly increasing work load. This impact will occur throughout the ADP environment including service support areas such as job preparation, tape library storage and retrieval, and quality review.

2.2.5.5 Processing Work Load by Functional Areas. The work load projections provided to the study team were also examined to establish trends based on functional areas of activity. As with the examination of current system utilization, the objective was to establish trends as related to production activity and in turn to relate the projected work load to specific data systems. Table 4 provides a summary of the most significant growth factors related to the DMIS/P line item and in turn to data systems. As would be expected, the most significant growth shown in the projection is related to the increased utilization of the SACARTS and DARIC systems in support of 1:50,000 scale line maps. The significant growth in this area also provided additional impetus to a more detailed examination of the SACARTS itself (Appendix A).

The total work load projected for the related product line items of 1:50,000 Line Maps, 1:50,000 Topographic Data Bases, and Digital Topographic Elevation Data was also extracted. The projected work load related to these products from 1976 through 1980 is shown in Table 5, and illustrates the sustained dominance of these products in the DMATC environment over the next five years. As should be expected, these same products reflect the most significant growth of data holdings.

2.2.6 Projected Data Storage Requirements. The projected growth of data storage requirements was initially based on a survey conducted by the DMATC Data Base Administrator. Data obtained through the survey was supplemented by briefings and interview. Size estimates for various files varied from detailed word count to "reels of tape" for which no data content estimate was available. To a large extent, however, active files reflected actual data content expressed in words or characters, while files expressed only in terms of reels of tape were of an archival nature.

Data file growth factors were also obtained during the survey and verified during the study period. Typically, growth rates associated with resource management files such as DMIS/P, TDLS, and ARAPS remained relatively low. ACRES related files reflected a higher rate of growth but total file size was still relatively small. The most significant growth relates to the direct production of topographic products in which UNAMACE master tapes, SACARTS related files (GIST Level IV), and DTIB elevation and planimetry tapes constitute a growing source data file.

The size of data holdings projected to FY 79 and FY 83 for the primary files addressed in the study is summarized in Table 6. As noted in the summary,

Table 4. Significant Changes in UNIVAC 1108 CPU Utilization[1] (FY 76 to FY 77)

Line Item	Includes	FY 76 CPU Hours	Change (Hours)	Percent Change
3AA 1:50,000 Line Maps	SACARTS/ DARIC	379	+2,423	+639
3EA Dig. Topo. Elevation Data	DTED	1,004	- 114	- 11
3GE DOD Library of Maps	TDLS	246	+ 62	+ 25
3GG Analysis and Evaluation	PMS	188	+ 26	+ 14
3GJ Procurement of Maps and Data	ACRES	240	- 24	- 10
4FD Geodetic and Gravity Studies		150	+ 125	+ 83
4FF Satellite Geophysics	CELESTE	297	+ 91	+ 31
4FG Special Surveys		140	+ 35	+ 25
7BA Mission ADP Support		185	+ 349	+189
7BC Product/Technique Development		384	- 78	- 20
7BD Program Management Info. Sys.	DMIS/P ARAPS	515	- 293	- 57
7BE Other Mission Support		268	- 86	- 32
Support For DMAHC		120	437	+364

[1]As of January 14, 1977

Table 5. Major Production Work Load[1]
(UNIVAC 1108 CPU Hours)

ITEM	FY 76	FY 77	FY 78	FY 79	FY 80
3AA 1:50,000 Line Maps	379	2,802	2,802	2,802	3,840
3 AC 1:50,000 Topo. Data Base	1,039	1,038	1,038	1,038	----------
3EA Dig. Topo. Elevation Data	1,004	890	1,172	1,172	1,172
	2,422	4,730	5,012	5,012	5,012
Total Validated Work Load	5,395	10,114	10,880	11,217	11,145
Percent of Work Load	45	47	46	45	45

[1]As of January 14, 1977

Table 6. Data Base Growth Projection[1]

File	Current		FY 79 Projection		FY 83 Projection	
	Tapes	Words	Tapes	Words	Tapes	Words
TDLS	50	72M	50	72M	50	72M
SACARTS	200		1,700		3,700	
DARIC	2,550		3,927		5,763	
DMIS/P	20	4M	20	4M	20	4M
DTIB						
DTED	1,100		1,800		2,500	
Unverified						
Planar	6,000		1,200		1,200	
Planimetry	80		1,580		3,580	
PMS	2	2M	2	2M	2	2M
ARAPS	4	7M	4	7M	4	7M
AMRR	Planned					
CELESTE	300	33M 98M(77)	480	(98M)	720	(98M)
ACRES and related files	16	34M	40	37M	72	41M
Totals	10,322	152M	10,803	220M	17,411	224M

[1]As of January 14, 1977

active files indicate data holdings increasing from a current figure of 152 million words to approximately 224 million words in 1983. The total growth of tapes associated with the primary files increases from approximately 10,000 at present to 17,411 in 1983. This volume of tape holdings includes a projected consolidation of unverified Planar tapes during the period from 6,000 reels to 1,200 reels.

Since the files of primary interest to this study reflect only a portion of DMATC magnetic tape holdings, a second summary was made which includes identified holdings in other files (Table 7), and holdings of the Department of Computer Services Tape Library. This summary also includes an estimate of tapes used as intermediate temporary storage in the production cycle. No growth was projected for this later group based on a management objective of maintaining a minimum number of tapes in this status. The summary projection of total tape holdings as shown in Table 7 reflects an estimated growth from an approximate level of 25,000 tapes to 35,000 tapes in 1983. In view of the rapidly expanding data processing requirement previously addressed, the data holding projection is considered conservative and will require a continuous management effort to assure minimum retention of magnetic tapes used for temporary storage.

2.2.6.1 Projected Source Data Base Holdings. Projected data holdings constituting a source/product data base are estimated to approach 17,000 reels of magnetic tape by 1983. This data primarily consists of UNAMACE master files, SACARTS-generated master files, and holdings in the DTIB consisting of both planimetric and elevation data compacted in DMA standard format. If it is assumed that these holdings will consist of consolidated data stored on 1600 bpi, 9-track tapes, and that the tapes are utilized at 75 percent of capacity, the data could represent a total holding of 4.7×10^{12} bits. The tape utilization figure of 75 percent is recognized as significantly exceeding the widely used commercial utilization factor which is normally 20-25 percent of capacity. The 75 percent factor is based on the assumption that cartographic file contents would be compacted for storage under the DTIB.

Projected holdings of this magnitude demonstrate the need to plan for improved methods of mass data storage. Since the holdings will consist of data which is primarily archival in nature, the choice of storage media will depend largely on the projected requirement for access and update frequencies. Efforts to verify and consolidate these holdings should be continued along with a further definition of projected utilization and response time requirements. These continuously refined estimates will provide a more stable requirements base on which an optimized solution addressing this expected mass storage requirement can be addressed.

2.2.6.2 Online Storage Requirements. One of the study objectives was to examine online storage requirements and provide a definition of resident and temporary online storage requirements. Resident online storage is generally described as storage space which can be accessed without manual intervention. Storage such as solid state memory devices, drums, fixed-mount disks, and fully automated mass storage systems fit this definition. Storage devices such as removable disk packs or floppy disks may be either resident or temporary depending on their use at any given time. Magnetic tapes are

Table 7. Projected Magnetic Tape Holdings

Item	Current	FY 79	FY 83
Data Files			
TDLS, DMIS/P	392	596	868
PMS, ARAPS			
ACRES, AMRR			
CELESTE			
Others	631	636	642
Source/Product DB			
DARIC			
SACARTS	9,930	10,207*	16,743
DTIB			
Others			
D/CS Tape Lib. etc.	4,000	5,200	6,800
Scratch and Work	10,000	10,000	10,000
Totals	24,953	26,629	35,053

*Includes consolidation of 6,000 unverified Planar tapes.

considered temporary online storage only since tapes are usually associated with a manual intervention.

Evaluation of online storage requirements is to a large extent a relative assessment depending on frequency of access, data volatility, and/or required response time. Using these criteria, the existing resident online storage of the DMATC consists of storage required for Catalogue Files and Scratch File Space. Excluding drum storage, these files represent 140 million words of disk storage for the Site I system. In terms of data files, files aggregating approximately 84 million words are maintained on removable disk and constitute the existing temporary online storage requirements.

This assessment of online storage requirements, either resident or temporary, is not considered adequate for future planning purposes. It is considered valid, however, in terms of the response time requirements identified by DMATC. In only two cases were required response times of less than 24 hours identified, these exceptions being CELESTE and ACRES. In our view the identified response times reflect only what the file managers expect in terms of turn around. A more realistic assessment should be based on production throughput objectives.

Potentially all the active data files should be considered as candidates for online storage. Based on current projections, data files in this category will reach a minimum level of 224 million words by 1983. This projection is over and above online storage requirements for catalogue file and scratch space which currently require approximately 200 million words of disk storage. Specific identification of candidates for resident versus temporary storage should reflect access frequency. A move toward integrated data management will also carry an inherent requirement for resident online storage.

The continued definition of online storage requirements should be addressed as an open topic within the DMATC environment as system improvements are implemented.

2.2.7 <u>Projected Restructuring of Data Files.</u> During the assessment of data holdings it was recognized that a broad range of opportunities exist which could improve the utility of existing data files held at DMATC. With few exceptions the files described in Section 2 are independent functionally-oriented files. Interfaces between interacting files are not in general automated and require manual extraction and preparation. None of the current files are maintained online, and data entry is primarily accomplished by means of punch cards or card images on tape. The only data system supporting online query is the DADMS.

Examination of the individual files has been initiated under the guidance of the Data Base Administrator. Efforts to date have recognized the interrelationships between many of the files by examining the function and role of the files. This approach is useful in identifying clusters of related files which, depending on frequency of use, may warrant integration, consolidation, or development of automated interface programs. An example of such interrelationships is readily apparent in the various directories or files (Fact

Files) which contain information regarding published/or in-production topographic maps. Some of the concerned files in this area are:

- DADMS--containing depot inventory and distribution files
- ARAPS--identifying user requirements and production status
- PMS--identifying currency and accuracy assessments
- DMIS/P--identifying personnel resources and schedules for current production
- REPROMAT--identifying reproduction materials

Similar clusters of files are readily identifiable in areas addressing the availability of source materials to support new production (TDLS, Analytical Triangulation, CELESTE, Geodetic Control, photo coverage files) while still another cluster includes files of processed data which could support new production requirements (DTIB, DTED, DTIB Planimetry, UNAMACE Masters, SACARTS).

In general all of the files fall in two broad categories which support management and direct production. In turn each of these groups can be further categorized by function:

- Management
  - -- Planning and Control
  - -- Resource Management
  - -- Source Directories

- Production
  - -- Operational files of processed data
  - -- Preprocessed source data
  - -- Product data bases

Motivation for examining any individual file or cluster of related files depends on a variety of factors such as frequency of use, stability of the file, how the files interface with the overall production system, and the volume of traffic. These same factors also affect the choice of how a selected file or set of files should be supported in an ADP sense to optimize the files utility in supporting the production effort. This can be best illustrated by some specific examples.

The DMIS/P is primarily concerned with the utilization of personnel resources and the monitoring of production schedules. DMIS/P is cycled on a weekly basis with a high level of activity associated with some 3,50C card inputs per week. Turn around times associated with the 15 individual files vary from 1 to 7 days which indicates that the DMIS/P could always be a week behind, diluting the usefulness of the system as a management resource control tool.

If a more responsive system is required, alternates available range from performing all batch processing in a pre-established time block under sophisticated software control, to placement of the DMIS/P on a fully online status

supported by interactive online update/query capability. In either case any improvements should be based on the role DMIS/P is intended to play in DMATC production and management's requirements for more responsive status reporting.

A second illustration of potential file improvement was apparent during initial presentations on the TDLS. The structure of this system was observed to allow addition of an online query capability for exploiting the file for support of production planning. The low query level projected, however, (35 to 50 queries per month) would not justify the additional costs of adding remote query terminals or additional online storage for the system.

These examples are cited to illustrate the broad range of alternatives that can be addressed relative to each file. Only through a comprehensive analysis of major file groups can a pattern emerge which provides a balanced recommendation.

2.2.7.1 Initial File Assessment. Restructuring DMATC data files to provide improved support of production is a complex and long-term task. The range of restructuring can vary from the development of relatively simple file interfaces to the consolidation of closely related subsets to the total integration of all files under a comprehensive Data Base Management System. The impact of restructuring similarly varies from relatively small software efforts to a complete revision in the concept of ADP support operations which are inherent in a totally integrated DBMS which could involve:

- Operation in an online, time-share mode
- Hardware to support the DBMS

One approach to initiating the restructuring of data files would be initiation of a study of high-activity files to identify candidates for possible integration or interfacing. The selection of candidate systems should be based on file activity, their general relationship to each other, and their importance to the production planning or support cycle. Typical groups of files might include:

- Map coverage files (DADMS, ARAPS, PMS, DMIS/P, REPROMAT)
- Photo Coverage Files
- Geodetic Control Files (Geodetic Control Points, Analytic Triangulation)

The initial assessment could examine specific fact files on a detailed basis to further define both short-term and long-term improvement requirements as well as an optimal data management structure for the files examined. This analysis would then provide a starting point for the systematic evaluation and upgrading of fact files into systems more fully integrated into the production planning or execution process.

2.2.7.2 Interaction with Standards. The move toward integrated data management and the use of data standards are mutually dependent and should be pursued as parallel efforts. The development of standards can be highly effective in facilitating file integration and interfacing, reducing redundancy in files, and reducing the data processing work load associated with format conversions. Conversely, the existence of a high level of data integration will inherently impose standards for data element definition, programming practices, program identification, and interface requirements for new files. When applied and considered during the development cycle of new equipment acquisitions, standards will minimize the requirements for reformatting programs and expedite the full integration of these acquisitions into full production use.

The extent to which a fully integrated data management concept is pursued is integral with long-range system architecture options. As will be noted in the following section, long-term objectives of the Pilot Digital Operations Plan (PDOP), assume a highly integrated data management capability structured under the control of a DBMS. On a short-term basis a major objective of file assessment should be a file-by-file analysis directed toward improved responsiveness of the specific data file being addressed. Whether such files should be integrated as a system, and whether or not they should be structured under DBMS should be assessed in parallel with the file analysis.

2.2.8 Long-Range Projections--Pilot Digital Operations Plan. Concepts relating to long-term objectives of the Defense Mapping Agency are being addressed within the context of a DMA production posture operating in a completely digital mode. The purpose of the Pilot Digital Operations program is to establish the technical and operational feasibility and the practicality (cost effectiveness) of performing each production operation in a wholly digital mode. The concept deals with the post-1985 time frame and is concerned primarily with digital image processing. As a secondary objective, the plan addresses the DMA posture involving digital processing in general, and in the context of a totally digital throughput production system. The relationship of this second conceptual objective is of particular interest to our study.

The PDOP primarily addresses the full scope of operations concerned with the manipulation and exploitation of digital imagery. The plan assumes that actions related to other forms of digital processing, such as that related to resource management and automated cartography, will progress at a rate which will support a totally digital production operation in the post-1985 period. The evolving DMATC data processing environment between now and 1985 should consequently be consistent with and supportive of the long-range DMA objective as conceptualized in the PDOP.

2.2.8.1 Systems Concept and Assumptions. The systems concept for Pilot Digital Operations is heavily oriented to the maintenance and exploitation of three primary data bases functioning in an online mode. The primary data bases are:

- Online Production Planning and Management System
- Digital Imagery and Parameters Data Base
- Products/Supplementary Source Data Base

The concept further addresses the role of each of these data bases through the entire production cycle. For our purposes it is not necessary to review that aspect of the plan. Of significance, however, are some of the assumptions included in the plan as they pertain to areas outside the domain of digital imagery. The assumptions are most pertinent to the Online Production Planning and Management System which is excluded from the scope of specific PDOP development objectives. Operations or functions excluded include the automated cartographic function not dealing with imagery, management information systems, and management functions.

Some of the characteristics of the Online Production Planning and Management System and the Product/Supplementary Source Data Base which are either explicit or can be inferred from the PDOP include:

- Online status
- Remote query capability
- Integrated data management
- Security
- Interactive display capability
- Interactive distributed work station capabilities
- Product independent data
- Electronic transmission capability

These characteristics are also evidenced in specific operations to be examined within the PDOP context. A summary of the more apparent characteristics is illustrated in Table 8. Since all of the objectives relate to digital image processing, only those with broader implications to this study have been identified.

2.2.8.2 Relationship to the System Architecture Study. In a strict sense, objectives of the PDOP cannot be construed as requirements since the purpose of the program is to ascertain the feasibility and practicality of the various function subsystem developments. As a growth objective, however, the PDOP establishes additional guidelines for system growth in the 1977 to 1983 time frame. These guidelines are of particular significance in the assessment of alternatives where selection of a course of action based solely

Table 8. Implied Technical Characteristics of Pilot Digital Operations Subsystems

PILOT DIGITAL OPERATIONS SUBSYSTEMS		CHARACTERISTICS: Network	Distributed Processing	Mass Storage	Online Data Base	Interactive Graphics	DBMS	Security
III.2	DMA Data Telecommunications	■	■					■
III.3	Automated Digital Screening and Assessment			■	■	■		
III.4	Digital Image Data Base Storage/Retrieval/Security/ Protection			■	■		■	■
III.5	Digital Image Data Base Query/Update/Purge	■			■		■	
III.6	Change Detection Between Digital Imagery			■	■			
III.7	Structure of Subset of the Products and Supplementary Data Base			■	■		■	
III.8	Man/Machine Interface through Digital Displays		■			■		
III.9	Product Digital Image Evaluation	■	■		■	■	■	
III.12	Digital Control Base				■		■	
III.14	Automated Change Detection (Image to Data Base)			■			■	
III.15	Compression of Digital Image Data			■				
III.30	Preparation of Press Plates from Digital Imagery	■	■		■			
III.31	Direct Printing from Digital Data	■	■					

on current needs may or may not be consistent with long-term objectives of the PDOP. To cite one such example, under the existing SACARTS concept, digital recordings of cartographic features are maintained in a product-dependent system using Cartesian coordinates. This approach is well suited to the production of 1:50,000 scale topographic maps, provides an efficient data storage mechanism, and avoids the computational burdens of high-volume geographic coordinate transformations. Conversely, the PDOP addresses a product-independent data structure using geographic coordinates. This dichotomy is illustrative of differences in short-term and long-term objectives which will have to be addressed at some point in time if the long-term objectives of the PDOP are to be achieved.

3. CURRENT RESEARCH AND DEVELOPMENT PROGRAMS

This section addresses three areas of developing technology of direct interest to the DMA for their application to the cartographic production process. Consideration of these programs for inclusion in the DMATC system architecture study is therefore appropriate. Two of these areas have potential for improving the data capture cycle supporting topographic data production: raster scanner/plotter technology and photogrammetric postprocessing. The third area concerns use of associate array processors to support the data processing function. The three areas are essentially independent topics both in terms of their status and their potential impact on the DMATC system architecture. The three topics covered in this section are:

- The Raster Scanner/Plotter
- The Integrated Photogrammetric Instrumentation Network (IPIN)
- Associative Array Processor

3.1 Raster Scanner/Plotter. Introduction of a raster type device to support data extraction and plotting is scheduled for use in the DMATC cartographic production system in mid-1977. This precision device is programmed for interface with the SACARTS process to take advantage of raster technology in cartographic production.

3.1.1 System Functions. The major desirable characteristics of raster technology address two principal functions supporting the SACARTS process:

- Scanning Function--A black and white scanning device serves as a data capture device using black and white manuscripts. The device can scan the equivalent of two Joint Operational Graphic (JOG) manuscripts in 15 minutes and represents an extremely high rate of data capture. In the DMATC production environment the device is planned for the digitization of data for input to the GIST software system either from reproduction material or from photogrammetrically generated contour manuscripts which have been edited for "cartographic expression." The scanning function requires off-line computer processing support for the conversion of raster formatted data to a lineal (vector) format. This function also requires an edit operation for the entry of feature classification headers.

- Plotting Function--The plotting function provides a precision plot on reproduction material as an alternative to the scribe coat color separation process. Plotter output is high and plot time remains constant regardless of information content. The plotter can plot the equivalent of two JOG's in 15 minutes. The plotter function requires data input in raster form. Reformatting of data from the SACARTS data bank (GIST Level 4, 6, or 7) requires off-line processing for the conversion from lineal (vector) to raster format.

3.1.2 Data Processing Impact. The scanner/plotter scheduled for implementation in the SACARTS process poses a significant data processing work load

on the UNIVAC 1108. Due to equipment delivery problems, the original support software developed for the system did not undergo a full development cycle. Based on initial tests of the system at the DMAAC, it was determined that approximately 20 hours of CPU time would be required for the lineal-to-raster conversion process for a typical map overlay. Since normal production requires or involves 5 overlays, the total requirement was estimated at 100 hours of CPU time per map. Based on a production schedule of 250 map sheets per year a resulting burden of some 25,000 CPU hours per year has been cited. This, together with an estimated 10 hours per chart for the raster to vector conversion process, suggests a total requirement approaching 27,500 hours per year.

Software supporting the scanner/plotter is currently undergoing improvement under contract to the Rome Air Development Center. Informal estimates regarding these improvements reflect a target representing a 75 percent reduction in CPU processing time. On a production basis supporting 250 map sheets per year, this work load would still exceed the entire DMATC processing time accrued on the Center's two UNIVAC 1108's during FY 76.

Improvements to the DMA scanner/plotter software addressing the vector-to-raster process and the conversion of the RADC raster conversion software and the ETL CASPP software to the UNIVAC 1108 will provide an initial basis for utilization of the scanner/plotter in conjunction with the SACARTS. Use of these programs will severely tax the available UNIVAC 1108 resources as the number of map sheets increases to the production goal of 250 sheets. The rate at which this goal can be achieved will also depend on the extent of the editing required to support the raster-to-vector conversion and the entry of data into the GISTS structure.

3.1.3 Impact on SACARTS Production. The most dramatic effect of introducing raster technology in the SACARTS environment is expected in the reduction of the production pipeline. Depending on the extent of the edit requirement, the time required to create a GIST Level 4 master record should be drastically reduced.

Although a significant increase in data processing requirements is projected, some trade-off will occur in terms of the current processing associated with manual digitizing operations and data preparation for lineal plotters. Additional trade-offs will accrue in the form of labor costs during the digitization and color separation phases. All of these factors will contribute to the overall assessment of the cost effectiveness of the DMA scanner/plotter in the production environment. The relative costs of manually produced maps versus those supported by the scanner/plotter should be subjected to a detailed cost analysis once the system has been integrated into the SACARTS production system. This analysis can then provide additional input to the assessment of improvements to effectively exploit raster technology on a full production basis. Although the current examination of the scanner/plotter points to data processing capabilities as the major constraint, actual implementation in the SACARTS environment may indicate the existence of other choke points, such as editing and feature header entry, that may lead to the identification of alternatives that provide a balance of data processing requirements with overall production throughput.

3.1.4 Consideration of Alternatives. Data processing associated with both the raster-to-vector and vector-to-raster functions involves the manipulation of massive quantities of data. Extensive software development has been addressed in a research environment at both RADC and the U.S. Army Engineering Topographic Laboratories. Some of these programs and the associated computers include:

Raster-to-Lineal Conversion

Computer-Assisted Scanning Techniques--RADC PDP 9
Computer-Assisted Color Scanning--RADC HIS 635
Raster-to-Lineal Software Conversion--RADC PDP 11/45
Raster-to-Lineal Software Conversion--RADC H-6180
Cartographic Scanner/Plotter Program (CASPP)--ETL IBM 360, CDC 6400

Lineal-to-Raster Conversion

Format Conversion Software--RADC HIS-635
Raster Imaging Software--RADC H-6180
CASPP--ETL IBM 360, CDC 6400
DMA Raster Plotter Software--RADC UNIVAC 1108

Two of these programs, the RADC Raster-to-Lineal Software System and the ETL CASPP system are currently being converted to the UNIVAC 1108 for use by the DMA. The CASPP system, adapted to use the DMATC scanner magnetic tape format as input and the GIST CALMA/NOVA magnetic tape format as output, will provide a raster conversion capability consistent with current SACARTS data formats.

The existence of similar programs on different computing systems (i.e., Raster-to-Lineal on PDP 11/45, H-6180 and UNIVAC 1108 and Lineal-to-Raster on H-6180 and UNIVAC 1108) provides a basis for a relative performance assessment. In addition, the demonstration of the raster-to-vector, and vector-to-raster format on the Goodyear STARAN array processor provides a further basis for examining alternatives.

It is readily apparent that alternatives exist in terms of the processor (sequential or array), size of the processor (large scale or minicomputer), and programming language. The capabilities cited also represent differing approaches to solving specific problems. These differences may occur in the way data is formatted, how general functions such as the sort and merge functions are handled, or the manner in which specific cartographic problems are solved.

For example the technique used to generate line weights suitable for portrayal of a road may consist of an expansion of a centerline with overlapping exposures (DMA scanner/plotter) or may be a double cased centerline with area fill (raster imaging). The same problem addressed on the STARAN uses still another innovative approach. While some basis for comparison is now becoming available, existing data reflects a variety of data volume, data resolution, program objectives (for example, level and type of symbolization), and techniques. A carefully designed comparative analysis should insure use of consistent parameters and test constraints to better define the optimal hardware, programming language, and software technique.

3.1.5 Potential Use of the ETL STARAN Processor. Implementation of cartographic programs developed by Goodyear Aerospace Corporation on the ETL STARAN processor could hold potential for supporting DMATC production. An accelerated program to install these programs on the ETL processor would provide the opportunity for further development of the cartographic applications software using DMATC production data. The full implications of such a quasi-production support program will require a detailed evaluation, but a first step in any event is the implementation of the Goodyear programs and their evaluation using high volume, high resolution data. It should also be noted that problems associated with data edit and entry of feature header data would still require UNIVAC 1108 processing for creation of GIST Level 4 master records and significantly offset the high-speed processing capabilities of the STARAN.

3.1.6 Conclusions. Completion of the raster scanner/plotter and the planned improvement in associated software programs will provide a basic capability to support both a data capture and a finishing plotter function for the DMATC.

The rate at which the scanner/plotter will be able to assume a major production role, i.e., support of 500 map sheets per year, is highly dependent on other system capabilities. The most critical of these are devices to perform edit and feature header entry functions.

The overall impact of the raster finishing plotter/scanner is largely unknown at this time for the following reasons:

- System performance of the scanner/plotter will not be fully demonstrated until mid-1977.
- The DIODE System, which is programmed to perform edit and feature header entry functions, is still in the implementation phase. Throughput capacity of the DIODE will be uncertain until full implementation and demonstration
- The full implication of performing edit and feature header entry functions on linealized scanner/plotter data using other data entry devices (DPC, COMPUGRID, CALMA) is largely unknown. Utilization of these devices for edit purposes will still require processing through GIST Levels I, II, III, and IV.
- The extent and nature of transformation programs such as DIODE to GIST format is unknown.

Because of the above factors an extended period of operational testing and integration into a full production posture must be anticipated.

A research and development effort should be considered to address alternatives to the data processing problems associated with use of the scanner/plotter in a production environment. This effort could include an assessment of processing options as well as interface options from a production throughput point of

view. The assessment should be based on carefully designed control conditions--data volume, resolution, function, etc.

Potential use of processing capabilities at RADC and particularly the existing interface between the H-6180 and the STARAN should be considered for this purpose. The interface of these processors under control of the MULTICS operating system provides a unique opportunity to assess relative system performance and module efficiency.

3.2 Integrated Photogrammetric Instrumentation Network (IPIN). The IPIN system currently being designed addresses the conversion of data from automatic photogrammetric compilation equipment for the production of digital terrain elevation data. The system will provide a means of converting, editing, and formatting photogrammetrically-produced data close to the point of data capture and at a rate consistent with the throughput capacity of the photogrammetric systems. The system will perform the variety of post-processing functions now performed in a batch mode on the UNIVAC 1108 processor.

The primary motivation for implementation of an IPIN type system is to provide a post-processing cycle consistent with the high data volume output of data collection devices such as the AS-11BX and clusters of AS-11 type analytical stereocompilers and stereocomparators ("pooled minis"). While the system capacity is tailored to meet the photogrammetric output of the DMAAC, the conceptual approach to performing the post-processing functions provides an excellent example of how a remote processing system can be implemented to provide a finished product consistent with the rate at which data is collected.

A simplified illustration of the IPIN is shown in Figure 6 and is useful in describing the functions performed by components of the system.

Input data is received from the "pooled-mini" systems and the AS-11BX/ACE.

The "pooled-mini" systems function as data collection devices. They also provide control data for the AS-11BX/ACE. The systems provide digital geomorphological records, i.e., drainage patterns and ridge lines which will be used to support terrain modeling during the matrix interpolation process. The "pooled-mini" systems provide digital output in geographic coordinates. The AS-11BX/ACE is the primary data collection device and outputs data in a model coordinate system. The output of this device (approximately 485 points per second) establishes the throughput requirements for the remainder of the system.

The System Processor collects and passes data from the collection subsystems to the remainder of the IPIN. The system processor also performs "on-the-fly" coordinate transformations for data input from the AS-11BX/ACE subsystems and converts data from model coordinates to geographic coordinates.

The File Manager controls data flow within the IPIN and directs data to temporary storage on magnetic disks or tape. Sufficient storage is provided to accommodate a system flow capable of supporting 10 models of ACE data. Data stored on disk is formatted in small sorted blocks, based on the one degree file structure prescribed in the "DMA Standard for Digital Terrain Elevation Data File." The File Manager grid indexes the data prior to storage on disk.

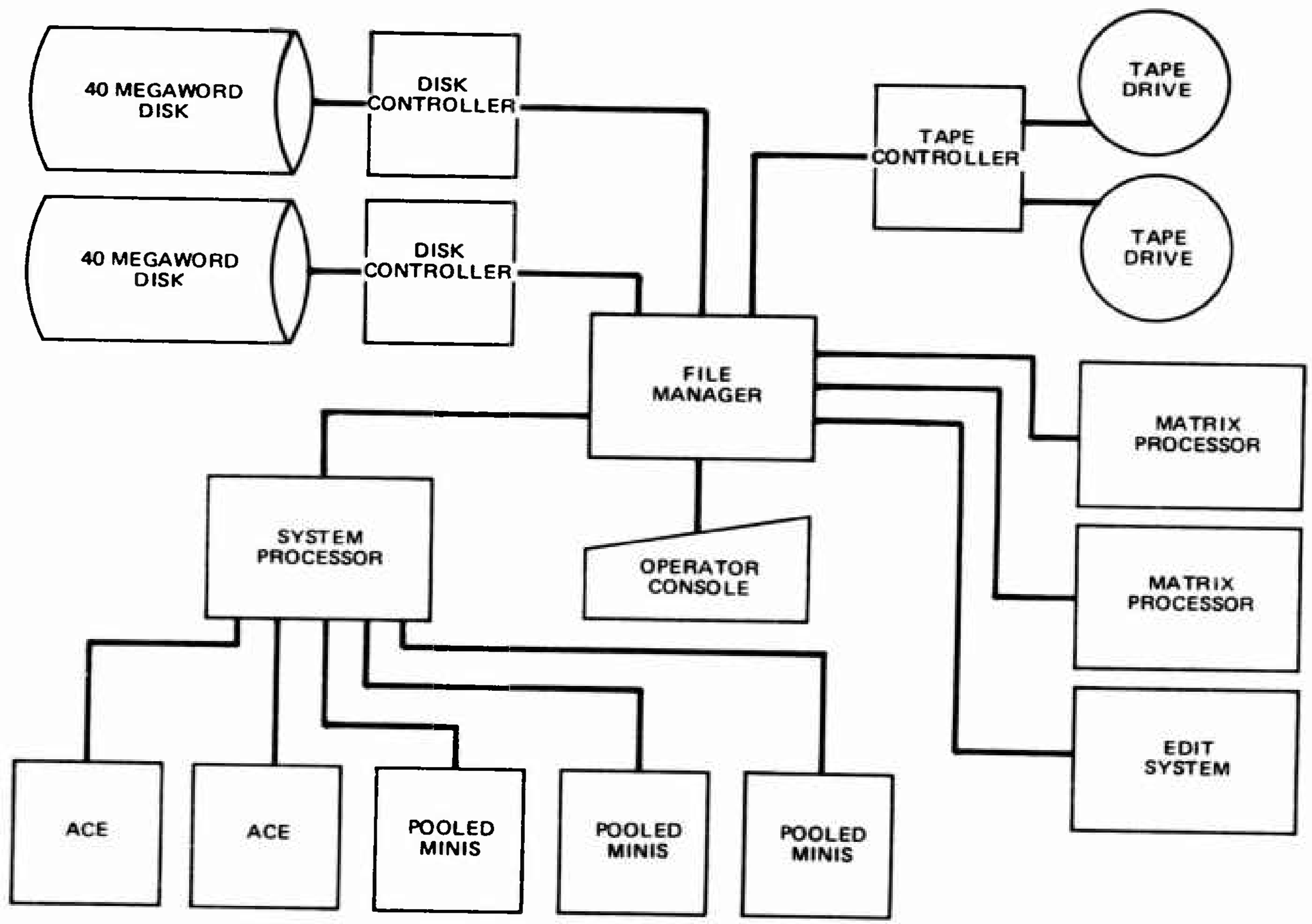


Figure 6. IPIN System Configuration

In other words, contiguous data on the earth's surface is similarly mapped to contiguous areas on disk. The primary data association element is a 3 arc minute by 3 arc minute data cell, which presumably corresponds to one track-cylinder on the moving head disk (each data cell would be composed of approximately 60 x 60 = 3600 elevation points). Similarly, these data cells are mapped in a directly addressable fashion to produce the basic data file which corresponds to 1 arc degree times 1 arc degree grid on the earth's surface. Thus, the basic data file consists of 400 (data cells) track-cylinders of information or approximately 1,440,000 elevation samples.

The structure of the disk file was designed on the basis of minimizing disk I/O time which turns out to be the most time-critical process in the system. Since the head movement of the disk is the most time-consuming component of the disk I/O time, two disk drives are used. One disk holds the data while the file control directories are maintained on the other disk. Thus data cells can be optionally retrieved from the data disk by placing contiguous data cells on adjacent track-cylinders. Displacement of the disk head is minimized as adjacent data cells are processed. The second disk allows for the avoidance of long excursions of the head of the data disk for periodic directory accesses.

After the ground data has been grid indexed and stored on disk, the Matrix Processor retrieves data cells via the File Manager for terrain matrix processing. The matrix processor performs an elevation interpolation to convert the data from the photogrammetric collection devices to a DMA standard matrix format. Since the data has been presorted, it is estimated that the time to process a complete data file will be on the order of 20 minutes. On the UNIVAC 1108 the same process now takes about 4 hours.

The proofing and revision processes are accomplished at the Edit System. There are basically two kinds of errors: (1) gross errors in which elevation samples were simply not recorded by the analytical stereoplotters; and (2) resolution of drains and contours and insertion of geomorphological data. The gross errors can be easily spotted by displaying the data and detecting "holes" or "gaps" in the elevation data. This type of error can be easily identified at the Edit Station. Type (2) editing is accomplished by capturing the geomorphological data (drains and spot elevations) at the Pooled Minis. The type (1) and (2) revisions are sorted back into the data file. The geomorphological data is lineally formatted for the revision cycle. When such corrections have been made and processed through the IPIN, the final edit will be made in the edit subsystem and the completed file transmitted to the UNIVAC 1108 for integration in the DTED data base.

The operating system envisioned for the IPIN will be very simply structured. The system will operate in a store-and-forward fashion, and will be essentially message driven. Each message will be accompanied with a destination code indicating the process to be performed. Priority queues for each message type will be maintained in each node processor in the network. Multiprogramming will not be needed by the node processors because all requests will be flushed out of the highest priority queue before processing other requests. Consequently, only the program that processes the highest priority queue will need to be in the node processor's memory. Further, since the disk I/O is being

performed by the File Manager, the other node processors can operate essentially in a CPU-bound mode (no I/O waiting). The average I/O traffic load is expected to be only about 8 megawords per hour through a single point on the network. The operating system will be transportable to any processing node on the network due to the modularity of the computers and the single-thread nature of the processing to be performed.

The system design will allow the proper selection and sequencing of functional operations such as matrix generation and editing appropriate to any specific data subset. Iterative processes can be continued as necessary to assure complete processing of a model prior to its release to the DTED data base.

3.2.1 Applicability of the IPIN to DMATC. From a conceptual point of view the IPIN is significant in that it addresses a recognized need to perform data edit and formatting functions as an integral part of data capture. By performing these functions on a time basis consistent with the rate of data capture, time lags associated with sequential passes through a batch-oriented central processor are eliminated. The approach also illustrates the manner in which data storage structure can maximize processor efficiency by minimizing the movement of data.

Consideration of an IPIN type system for the DMATC environment must recognize some basic considerations which could significantly alter the configuration and functional objectives of a DMATC-oriented IPIN.

- Capacity of the IPIN is tailored to match the output of three "pooled-mini" systems and two AS-11BX/ACE stereocompilers.

- The DMATC "pooled-mini" program addresses the clustering of the Center's AS-11A stereocompilers. The AS-11A is a manually operated stereocomparator capable of generating a graphic output in either profile or contour mode. Although the AS-11A did not provide digital recording, the "pooled-mini" approach will provide a digital record in either mode. Throughput capacity of the AS-11A's, even in a cluster configuration will remain well below other devices (Pooled AS-11B1, AS-11BX, ACE).

- Primary digital output from photogrammetric compilers at DMATC consists of profile recordings from the UNAMACE. Pending additional augmentation or replacement of the UNAMACE by equipment similar to the AS-11BX, the UNAMACE output would establish capacity requirements of a DMATC-oriented IPIN.

- The IPIN is oriented towards the production of digital terrain elevation data in matrix form, while DMATC's primary production requirements are directed toward the line map and the elevation contour.

Recognizing these inherent differences, it would be relatively easy to configure a DMATC-oriented IPIN tailored to match the output of the data capture devices. The retention of geomorphological features to serve as constraints during the terrain modeling function could also offer the opportunity to

functionally orient a system to produce both edited contours and terrain matrix data. Such an approach would directly support DMATC's production requirements involving line map production as well as produce terrain matrix data as a direct useful output. The approach would also allow the formatting of data under the SACARTS GIST structure for systematic retention of terrain data as an integral part of a topographic information data base.

3.2.2 Summary. An IPIN type system tailored to the DMATC environment would impact the current DMATC production process significantly:

- The production of contours and elevation matrix data would occur at a pace consistent with photogrammetric data capture capacity.

- Throughput time would be accelerated, since delays encountered in batch processing through a central processor would be avoided. These delays are incurred by data handling involved in processes of editing, plotting, contour generation, etc.

- Direct production of edited contours and terrain elevation matrix data early in the production cycle would affect the process by which contours are now corrected for geomorphological constraints in both the Department of Cartography (UNACOMP) and the Field Offices.

3.3 Associative Array Processor. The purpose of this section is to provide general background information, specific alternative uses of associative array processing technology for DMATC, and recommendations for implementation of the most plausible alternatives. The background information on associative array processors in general and the STARAN in specific, includes definitions, characteristics, advantages, disadvantages, and application areas. The application areas most relevant for DMATC operations will be identified as possible alternatives and the rationale behind the recommendations will be discussed.

3.3.1 Definition of Terms. Many computer applications require identical operations on multiple streams of data. Most present day computers (monoprocessors) accomplish their tasks through the use of a single central processor unit (CPU) that is capable of addressing only one datum within any one stream per instruction cycle. When the number of data streams (or, alternatively, the number of data elements to be processed), becomes large, even the fastest of these computers (at speeds approaching several hundred nanoseconds per cycle) may require a completely unacceptable amount of time to complete its tasks. In light of these difficulties, considerable work has been directed toward more efficient techniques for "bulk" processing. The results have emerged as multiprocessors, parallel processors, and associative processors.

The multiprocessor employs more than one CPU sharing a common memory and peripherals. Each CPU is capable of independently executing its instructions (or in responding to conditions set by other processors), and each CPU is basically similar to the CPU found in conventional monoprocessor configurations. Through appropriate interprocessor linkage, a multiprocessor is then able to either process more than one stream of data or is able to assign one CPU to each sequential segment of a program operating on one data stream (or both).

This sequencing is known as pipelining. However, despite the significantly improved throughput rates attainable with a multiprocessor configuration, the programming techniques remain basically the same--an entirely sequential instruction set for each CPU where each CPU may handle only a single data stream.

Parallel processors (or array processors), are characterized by the handling of multiple data streams with the processor for each stream controlled by a single master CPU. However, the relative locations within memory of each data stream are fixed and the data structure must be completely known in advance. Data addressing is by physical address (location in memory) only--just as in the case of the conventional monoprocessor.

The associative processor, as currently implemented (e.g., a STARAN) is characterized by its method of memory addressing as well as by its inherent parallel processing capabilities. As opposed to a strictly physical address, operands in an associative processor may be accessed or selected based upon the contents of the memory. This form of memory organization is called content-addressable memory, or CAM. The inherent power of such a memory organization lies in its ability to address all data "associated" with a given descriptor simultaneously and is limited only by the engineering practicalities of implementation.

The primary difference between an associative (i.e., involving CAM) and non-associative parallel processor may be demonstrated by a simple example. Given a memory containing a large data field in which a binary pattern of interest may (or may not) exist, it is desired to test for the existence of the selected binary pattern within the memory and to determine the location of the pattern (or patterns) so that the pattern may be altered. With a parallel processor, we would still have to search each memory location for a pattern match (even though a constant could be subtracted from all locations simultaneously, thereby allowing a test for binary 0's). Then for each location satisfying the match, a new value would be written, one location at a time. However, with CAM, all memory is simultaneously queried with one instruction and all locations containing the desired pattern are immediately identified by the contents of the memory response store (equal to one bit for each word). Further calculations may then be performed using only the words responding; the alteration is performed on all responding words by a single "store constant" command.

Figure 7 illustrates the main features required to implement an associative address operation. In Figure 7, the application may require that all cities with elevation greater than 500 feet above sea level be located. This would be accomplished by performing a greater-than-comparison search in the elevation field of the file. This search is done in parallel. To set up the search a Data Register is loaded with the elevation (500) used for comparison, a Mask Register is included to mask the data searched, a Word Select Register specifies the subset of words to be addressed, a Results Register collects the results of the search, a Match Indicator is used to indicate the number of matches, and a Multiple Match Resolver indicates the "topmost" matched word. A six-word example of a geography file is illustrated in Figure 7; it assumes

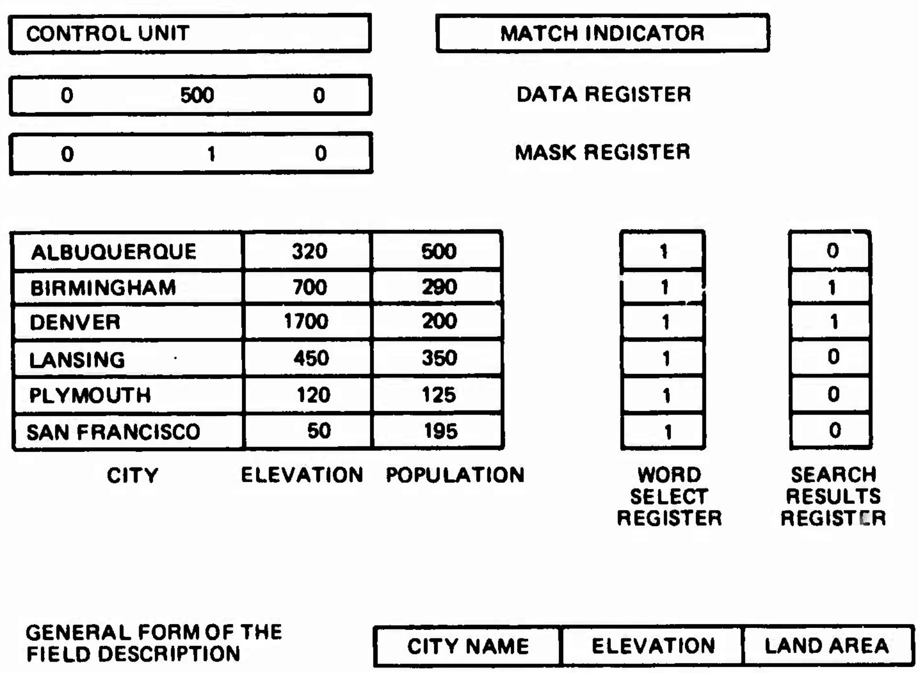


Figure 7. Associative Address Operation

that the search for elevation greater than 500 was performed on the entire file. All necessary registers and features are illustrated.

3.3.2 General Applications of AAP. In general, the array or parallel processing capability provides for fast, high-performance and real-time processing ability, while the associative processing characteristic provides for fast data searches and flexible data organization. Several areas of application appear quite well suited to associative and parallel processing. They provide cost-effective system augmentation in some cases (e.g., air traffic control and bulk filtering). In others, AAP's are very close to functional analogs of the physical system (e.g., data compression, communication multiplexing, and signal processing).

The use of AAP's appears promising in the elimination of critical bottlenecks in current general-purpose computer systems. Only very small dedicated associative memories are required and cost is really not a critical factor in the selection of these problems. Such problems may be encountered in the management of computer resources and might involve such items as protection mechanisms, resource allocation, and memory management. A major application area for which AAP's appear ideally suited, but due to cost factors have yet to prove themselves, is in data base management or large file searching and processing.

The application advantages of AAP's are:

- Low cost to produce--using LSI technology
- Provides a growth option--modular concept
- Increased throughput--using fast search capabilities
- Off-loading capability--can take some load off host, general purpose machine
- Reliability--due to modular concept

The general disadvantages of AAP's include:

- Complexity of program design--stems from the difficulty devising algorithms to best utilize the unique features
- Throughput degradation--may occur if single I/O channel is provided in either the AAP or host machine

Since the associative array processor (AAP) combines the technologies of other processors and is a prime concern of this study, we shall focus on it for the remainder of this section. Because the STARAN is the only AAP commercially available at the present time and is potentially available to the DMATC, we shall specifically direct our attention to it.

3.3.3 STARAN Characteristics. To accommodate the fact that AAP's are designed to be special purpose machines with various possible uses, and thus,

need to interface in a wide variety of ways, the STARAN system was divided into a standardized main frame and a custom interface unit (Figure 8). A variety of I/O options implemented in the custom interface unit include conventional direct memory access (DMA), buffered I/O (BIO) channels, external function channels (EXF), and parallel I/O (PIO).

The main frame consists of a conventionally addressed control memory for program storage and data buffering, a control logic unit for sequencing and decoding instructions from control memory, and from 1 to 32 modular associative processors. A key element of this system is the "main frame" memory providing content addressability and parallel processing capabilities. Each memory array is a multidimensional access memory matrix of 256 words by 256 bits (i.e., 65,536 bits) with parallel access to up to 256 bits at a time in either the word (horizontal or row) or bit (vertical or column) direction. Each array contains a 256 bit-serial processing element, each of which has an independent external device I/O path. Control signals generated by the control logic unit are fed to the processing elements in parallel and all processing elements execute the instruction simultaneously.

STARAN utilizes a PDP 11 as its sequential controller and has a sequential associative processor control unit which has a memory cycle about ten times faster than the PDP 11.

Although the STARAN's associative memory allows for multidimensional accessing, only one set of registers (X, Y, M) and processing logic are needed. The X-register is generally used to store temporary results. The Y-register effectively acts as the search-results register. It typically contains the results of search, arithmetic, and logic operations. The M-register is used to specify element activity. This register, in the bit-slice (vertical) mode, corresponds to a word-select register, and in the word-slice (horizontal) mode, to a mask register. The STARAN does not have a dedicated serial adder on a per word basis. Goodyear programs the X- and Y-registers, using the logical functions, to appear to have this capability. This cuts the cost of an array, but requires the facility for high-speed operation within the X-, Y-, M-register complex.

Each of the 256-word X 256-bit associative arrays includes a 256-bit resolution system. Resolution is always going on in each array, and the system interface is a 9-bit response output--8 bits giving the address of the first responder, and the ninth bit is the Inclusive OR of the response register.

The system I/O is not definable because each system is unique in the kind of I/O required. Typical options include DMA to a host computer, buffered I/O for peripherals, and communication through EXF and PIO channels into any of the arrays.

The assembly language APPLE--Associative Processor Procedural Language--has been developed for STARAN. Assemblers for APPLE, including macro assemblers, are available and are tailored for individual machine installations. Few I/O instructions are included in APPLE, since I/O is also customized for each installation. To our knowledge no high order language has been developed and implemented for the STARAN.

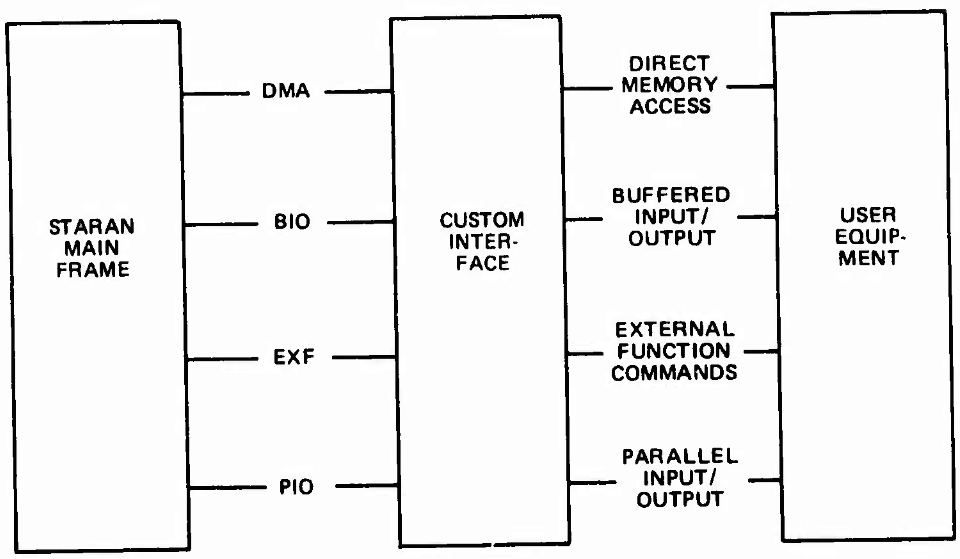


Figure 8. STARAN System

Thus far, three STARAN systems have been delivered: Rome Air Development Center, Johnson Space Center at NASA (Houston), and the U.S. Army Engineering Topographic Laboratory.

3.3.4 DMATC Applications for the STARAN. Research and development activities related to examination of associative processing technology have been pursued through DMA-sponsored programs at both Rome Air Development Center and the U.S. Army Engineering Topographic Laboratory. The thrust of these efforts has been directed toward the investigation of applications of interest to the DMA and the identification and development of support systems which would be required in a production environment. The major thrust of these efforts is summarized below.

3.3.4.1 Synopsis of Associative Array Processing of Raster Scanned Data for Automated Cartography. This project was funded jointly by the Defense Mapping Agency and the U.S. Geological Survey. The objective was to assess and show the feasibility of using the STARAN for raster processing including raster-to-lineal conversion. This processing entailed generating a map sheet with a variety of representative map symbols such as single lines, double lines, broken lines, railroads, area fill, and point symbols. The source data was raster scanned with a resolution of 4 mils and had a map image area of 19 x 22 inches. Also included as functions of such processing was to separate, as a function of line thickness, each feature class of a raster scanned, 4 class, feature separation and to vectorize a raster scanned contour sheet. The project designed and implemented the necessary software to produce the needed data for the production of a map using the STARAN and measured the timing performance of major software routines. The conclusions reached by the project team include:

- The STARAN could produce the basic types of map symbology in a timely manner (at an average rate of 125 seconds per symbol).

- The STARAN could efficiently perform vectorization and line separation tasks for cartographic work (at the average rates of 187 seconds for vectorization and 235 seconds for line separation).

While this project did show the feasibility of using the STARAN for map processing in conjunction with the ETL-IBM raster scanner/plotter for this one map, it has not established the cost effectiveness of the system in a production environment.

The demonstration illustrates the high processing speed of the STARAN. It also highlights the high I/O demands of the processor. The project report includes a recommended stand-alone STARAN configuration in which data compaction would be performed within the STARAN system. Use of a stand-alone configuration would facilitate a future integration of the STARAN in a production environment.

There is still much work to be done in modifying and adding to the present software to better present routines and provide for other necessary functions. The largest bottleneck occurs in the I/O area. To this end work must be done to reduce the amount of data passing in and out via compaction techniques and

with the use of higher speed devices. A smoother, more automated interface between the raster scanner/plotter and the STARAN should be developed, including provisions for editing and feature header entry.

3.3.4.2 Synopsis of Application of a Parallel Processing Computer in LACIE. The Large Area Crop Inventory Experiment (LACIE) is a joint investigation by NASA, USDA, and NOAA to determine the usefulness of computer-analyzed remotely-sensed data in crop forecasting on a global scale. The STARAN was used in the LACIE program for such pattern recognition functions as statistics collecting, iterative clustering, adaptive clustering, maximum likelihood classification, and mixture density classification. The LACIE algorithms are well suited to the AAP architecture because of inherent parallelism. The computation associated with each picture element (pixel) is the same for a given algorithm and can be implemented in a single instruction stream over the array of data points.

The STARAN is configured as a special purpose processor (i.e., slave) to an IBM 360/75 host machine. Originally all the work, including the pattern recognition, had been done on one of five IBM 360/75 computers. Therefore, all pattern recognition routines had to be redesigned and rewritten to make use of the features of an AAP.

The results and conclusions of the project team are very positive. The sponsors believe their success with an "essentially research-oriented system" will provide the basis of a production system for analyzing and employing multispectral data. This work will provide input for the next attempt toward a production system. Of importance and assistance will be the awareness of the performance evaluation dependencies between:

- Algorithm organization (ability to exploit parallelism)
- Number of data channels
- Number of classes/clusters
- Number of pixels per quantum of system work (job)
- Setup time for STARAN (formatting vector transfers to and from the STARAN)
- Data base retrieval rates

Again this is a promising application for the STARAN to DMATC operations. However, the transition from an experimental to an operational system is not a simple modification. Further work is needed in this area to produce the necessary evaluation and eventual transition.

3.3.4.3 Synopsis of An Implementation of a Data Management System on an Associative Processor. This project represents a research effort of Goodyear Aerospace Corporation to show the advantages of using an AAP coupled with a special head per track disk for Data Base Management System functions. A single level, hierarchical structure was chosen for the data. This technique

maintains each level of the hierarchy is a unique record type and associated with each record type are level codes indicating parentage of the particular record. A sample data base of 400,000 characters is structured into a four-level hierarchy (i.e., command, unit, sortie, and option records). The search technique used was entirely dependent on the physical layout of the data base on the disk. Due to the parallelism in the disk heads and the STARAN, the data is arranged in sectors consisting of 64 tracks where each track is composed of 256 bits. Thus, each sector contains one-fourth the STARAN array. If one assumes the entire data base is contained on one surface, two passes of the surface under the read/write heads will provide a search of the entire data base. The intent of this work is to have appropriate software to perform data definition, file create, interrogation, and update.

As of the last available report only the data definition and create modules were operational on a sample data base. A simplified version of the interrogation module was performing queries utilizing the Sigma 5 host computer to translate the query into a task list for the STARAN. Also operational were the change and delete options of the update module.

It was Goodyear's conclusion that this work provides great promise for the future of DBMS. They believe this is a viable way to handle the dual problems of providing rapid query response with easy update capability. Further work is necessary to prove cost-effectiveness. Simple performance evaluations can be made between the same data base queries with and without AAP assistance. While this technique may be efficient for small to moderate size data bases, the true problems do not actually arise until the size of the data base becomes very large. The transition from moderate to very large data base systems is not usually linear. This, too, is a promising area of application for DMATC, but much more work remains to provide operational capabilities.

In a separate RADC effort, Syracuse University has conducted preliminary investigations for the development of a Data Management System Strategy for possible application in a DMA environment.

3.3.4.4 Other Research and Development. Other research and development relevant to DMATC use of an associative array processor is discussed here.

3.3.4.4.1 Mass Storage Device. Under a RADC program a design of a one-half billion bit mass storage device has been completed. Fabrication of a bit access engineering prototype has been estimated to require a three-year development effort. Design of the device is generalized to allow use with parallel/array processors in general and is modular to allow expansion to a billion bit capacity. Projected costs of subsequent storage devices are estimated at about one million dollars. Implementation of a device fabrication program has not been initiated.

3.3.4.4.2 High-Order Language Development. Initial studies have been directed toward the design of a high-order language (HOL) for use with the STARAN. At the present time the STARAN must be programmed in a machine language unique to STARAN. Development of a high-order language and an associated compiler would allow applications programming in a commonly used

programming language such as FORTRAN. Availability of a HOL and compiler would significantly increase the ease of exploiting the STARAN capabilities in a production environment. Development of a HOL and compiler represents a considerable investment risk until a firm decision is reached to introduce a parallel/array processor in the production environment. Conversely, lead time requirements for compiler development could restrict full exploitation of the parallel/array processor capabilities until the HOL capability were available. A specification for a HOL has been delivered to RADC under an RADC/DMA sponsored R&D effort.

3.3.4.4.3 UNIVAC STARAN Interface Study. This study by the Georgia Institute of Technology addressed the feasibility and implications of interfacing a STARAN processor to a host UNIVAC 1108. In this configuration, the STARAN would function as a peripheral of the UNIVAC processor for the processing of applications programs. This study provides an initial assessment of interface implications and an alternative to a stand-alone configuration as conceptualized in the previously referenced Goodyear study.

The following summarizes the major points presented in this work:

- The study identified the two most effective (i.e., highest transfer rate) connection points. Due to bulk memory transfer rate constraints and additional internal logic needed, the fastest connection point, via an IOC port, was considered to be less cost effective than the slightly slower connection point via an ISI channel.

- Shared memory interface with the possibility of a corner turning memory addition was identified as a topic for further study. A queue-exchange system of buffering data was also recommended.

- DR11-B interface design was detailed including a $5,000 cost estimate. Further study was recognized as necessary in the areas of burst transfer impact on the STARAN and actual transfer speed.

- The UNIVAC software needed was defined and several algorithms presented along with the estimated cost in code and overhead.

- Data format conversion hardware unit was described and feasibility demonstrated.

- The data transfer rates of the two machines were studied and the conclusion that the UNIVAC 1108 cannot communicate with the STARAN fast enough to fully utilize the STARAN was reached.

As long as the average processing rate of the STARAN is greater than 10 milliseconds, the transfer rate of the UNIVAC machine would not cause any delays or idle time for the STARAN. Therefore, the feasibility and baseline constraints for a UNIVAC 1108/STARAN interface have been identified.

3.3.5 Recommendations Concerning AAP. It is undeniable that AAP's in general, and the STARAN in particular, provide attractive capabilities for

performing automated functions heretofor either too cumbersome to consider or totally unthinkable. In light of the DMATC production missions, the applications towards map production, image processing and pattern recognition, as well as data base management search capabilities, present great possibilities for further automation and efficiencies. The work thus far accomplished toward such ends shows promise, but has been of an experimental nature with little empirical evidence of cost effectiveness in production environments.

The efforts toward DMATC applications have been very fruitful in presenting the feasibility basis, providing the initial step, and pointing to specific areas for further work. Judging from the opinions formed in the work already cited, among the targeted areas for further works are:

- Design and production of parallel algorithms
- I/O procedures
- Distinguishing the roles of the sequential host and the AAP special purpose machine
- High-order language for the STARAN and better debugging tools
- Optimum word and array size (Goodyear has already announced a Module with 1024, 5120 or 9216 bits per word arrays)
- Resource allocation and utilization (e.g., control memory vs. array memory)

When the installation of an AAP is to be considered, various configurations should be evaluated. Since the STARAN (and any AAP) is a special-purpose computer, it will need a host computer connected in some manner. In an application such as map production or image processing, it may be desirable to provide a dedicated mini-host for stand-alone performance or a communication link to the UNIVAC 1108. This decision would depend on time, financial, I/O, and system degradation issues. For a data base management application, the decision between connecting the STARAN to the main frame system directly or through the back-end processor is interdependent with the decision to utilize back-end processing (section 4.1 of this report). Considerations of time, finances, I/O system degradation also arise as do issues of security and flexibility.

In conclusion, there is promise in AAP as applied to map production, pattern recognition, and data base management systems. Only a small amount of work has thus far been completed to make these applications efficient and reliable in a production environment. However, an AAP as well as other special purpose systems should be kept in mind while designing any new or upgraded system architectures. The key is to provide a modular configuration for maximum flexibility in order to add equipment or modify present modules in any way and specifically with the addition of a STARAN system. Such a modular configuration should include a flexible communication link between all system components so that such machinery as the STARAN may be connected to any one or several modules.

4. ADVANCED INFORMATION HANDLING TECHNOLOGY

This section addresses three subjects dealing with current developments which may be pertinent to advanced data management concepts at the DMATC. The first of these is a review and assessment of back-end data base management technology. This approach to integrated data base management may provide a viable option for future integration in an advanced system architecture. The second topic covered in this section addresses an updated assessment of current mass storage technologies and their future trends. The third section addresses security issues. During the course of the study the issues related to a multi-level security environment emerged as a major constraint in the adaptation of advanced ADP technology in the DMA environment. This topic has been introduced to focus attention on security as a growing architectural constraint and also to support the position that technological advances in both software and hardware will support a realistic solution to security problems in the near future.

4.1 Back-End Data Base Management

4.1.1 Introduction. Before addressing the topic of back-end data base management it may be helpful to briefly describe two general terms to establish a background reference:

- Data Base Management. A systematic approach to storing, updating, and retrieval of information stored as organized data items which are usually in the form of records in a data base.

- Data Base Management System (DBMS). A generalized software system assigned to manage the data base providing facilities for organization, access, and control. Organization facilities provided by DBMS software packages are aimed at representing the user's view of data in such a way that it can be accessed efficiently at a single storage location.
  - -- Access facilities of DBMS software packages provide the mechanism for storage, retrieval, and dissemination of data to and from the data base.
  - -- Control facilities included in the DBMS software packages help to maintain the data base with a high degree of integrity.

In brief, the DBMS is a software system which relieves the user from the task of routine data maintenance functions. The DBMS provides routine data storage and retrieval services to the user in a highly transparent manner. Most significantly itprovides program/data independence by making a single set of data, stored in a single location, available to many users for many programs.

A wide variety of DBMS's have been developed which reflect a wide range of function, complexity, and reliability. Traditionally, the DBMS was used on large mainframe processors where servicing the data management function competes with application programs for processor resources. More recently, DBMS's have been extended to a variety of minicomputers where the same contention for processor resources exist. The back-end data base management concept is a form of distributed processing in which the data base management

function and applications processing are distributed. The principal motivation for this functional distribution is the alignment of processing functions with machine capabilities and the maximization of application-oriented processor efficiency.

4.1.2 Definitions and Concepts. In essence, back-end data base management is a type of functionally distributed system architecture in which the data management routines of a host computer are off-loaded to a separate computer that is dedicated to operating those routines. The dedicated computer is called a back-end processor in analogy to front-end processors which manage communication routines and hardware. The expression "back-end processing" is synonymous with back-end data base management; for the sake of simplicity the acronym "BEP" will stand for back-end processing or back-end processor in the appropriate contexts.

A few qualifications may help to clarify this concept. First, a BEP is identified solely by its functional relationship to a host--the BEP performs the host's data base management operations. No other extrinsic characteristics (e.g., physical distance between host and BEP, relative computing power of host and BEP, etc.) are essential to the BEP concept. Second, the "application" software executed by a BEP is of necessity a data base management system (DBMS); however, nothing in the definition of BEP requires a particular kind of DBMS. Theoretically, any DBMS could be implemented on some kind of BEP, although it is unlikely that a given BEP will offer more than one DBMS. Finally, the data base on which the BEP operates need not be fully integrated, since the BEP concept is broad enough to encompass partially integrated data bases (at the price of increased system complexity and overhead).

Within this general context of data base management there can exist an unlimited variety of BEP systems, most of which share a few common elements: the BEP itself, some storage devices, and a control and data interface with a host (Figure 9). Data requests or data modification operations from a user application program or query are passed from the host to the BEP via the interface. These requests typically would be expressed as commands within a data management language (DML); each type of command would correspond to an executable routine (e.g., search, fetch, sort, insert, modify, delete, etc.) within the BEP.

Hardware for a BEP can range from microcomputer to large mainframe; the most cost-effective BEP's are likely to be mini-or micro-systems (since mainframes are usually designed for efficient numerical computation rather than data manipulation). The choice of a particular type of hardware will be determined primarily by the choice of a preferred data base management system (DBMS), since most computer manufacturers and vendors will provide their machines with only one DBMS. If it is decided to construct a DBMS for a BEP from scratch, then other considerations become important. For example, if the BEP were microprogrammable, then the most heavily used DML routines could be microcoded in order to decrease execution times. A large word size and available core space would eliminate the need for memory management routines with their associated overhead. A complete list of such desirable hardware features would depend entirely on the operational requirements placed on the BEP by the DBMS.

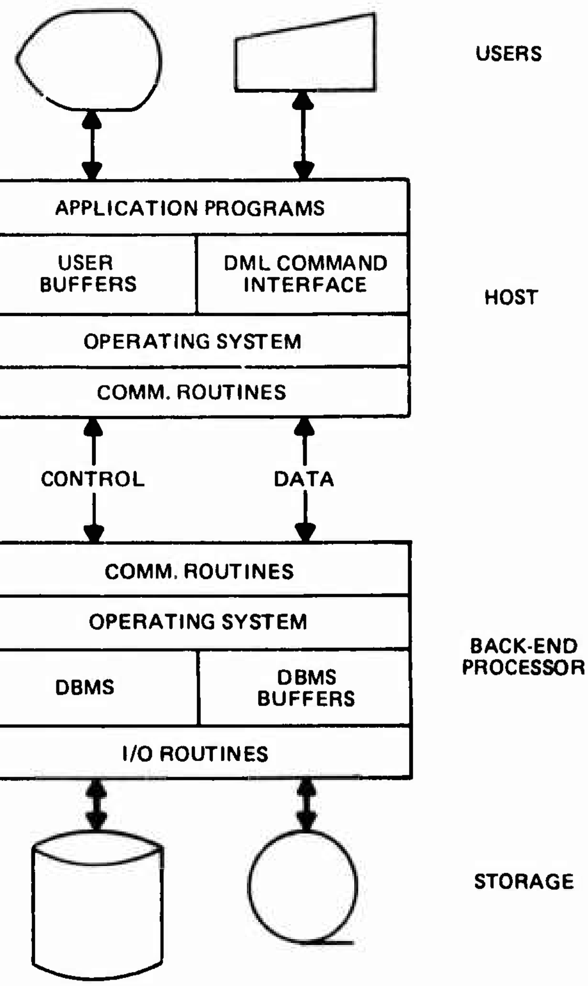


Figure 9. Hardware and Software Essentials of Back-End Processing

If time permits, a BEP can process DML requests from a host in serial order, but it is likely that neither the host nor the BEP can afford to wait for BEP I/O operations to complete in order to continue processing of a task. A BEP which is capable of interleaving and overlapping the processing and I/O for several DML requests concurrently is said to be multithreading (MT). A MT/BEP can also coordinate DML requests in a multiple host network.

A multiprocessing (MP) BEP has more than one instruction processing unit. Since it can operate on several DML requests simultaneously, a MP/BEP may have very large data throughput capacities, especially if each individual BEP in the system is also multithreading. MP/BEP's can be organized in one of three ways: a dedicated system in which each BEP accesses only one data base or one portion of a data base (Figure 10a); a distributed system in which each BEP can access all of the data base(s) (Figure 10b); or a hybrid system which is a combination of the other two alternatives.

Another important characteristic is the type of access to the BEP which is available to system users. One possibility is a host-dedicated BEP in which all communication between users and the BEP must pass first through the host; only the data base administrator would be allowed direct communication with the BEP. At the opposite extreme all users would communicate only with the BEP (i.e., there would be no host). Such a freestanding BEP would be in effect a stand-alone data computer with extensive data base management support but with limited arithmetic capabilities and language facilities. An intermediate alternative is the partially freestanding BEP. This type of architecture allows system users to issue commands to the BEP either directly or via application programs in the host (Figure 11).

One more attribute of BEP systems is significant. The variable portion of any DBMS is its application programs. If any of these programs are executed by the BEP (perhaps because the "pure" DBMS routines do not fully utilize the BEP's CPU) then the BEP is user programmable; otherwise, it is user non-programmable. Note that in this context the term "user" excludes the data base administrator, who is always allowed to program the BEP. Also, it is difficult to separate the functions of a user programmable BEP and a host; if a BEP routinely executes a wide variety of application programs, it begins to resemble a general purpose host.

4.1.3 <u>System Impact of Back-End Processing.</u> When a BEP is installed in a system, it will have significant impact on all hardware, software, and data base resources. While all of these resources are heavily interdependent, it will be convenient to discuss the BEP's effect on each of them separately.

<u>Hardware.</u> Since the BEP will contain all DBMS and file management routines, data base schema tables, and data base I/O buffers, the host will not need to maintain space for them in core. The resulting extra core can be used to increase partition sizes or the number of partitions in the host, or core could be reduced, thereby reducing the host's cost (if it is rented or leased).

The BEP itself will probably be a minicomputer with sufficient computing power to keep up with the rate at which the host issues DML requests, and

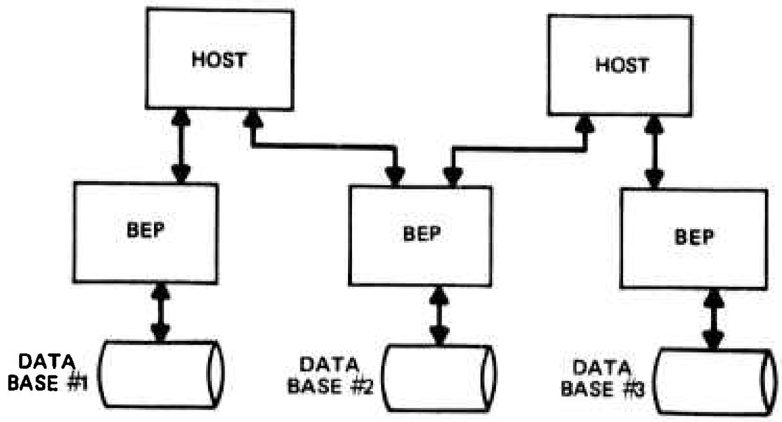


Figure 10a. Dedicated Multiprocessing BEP

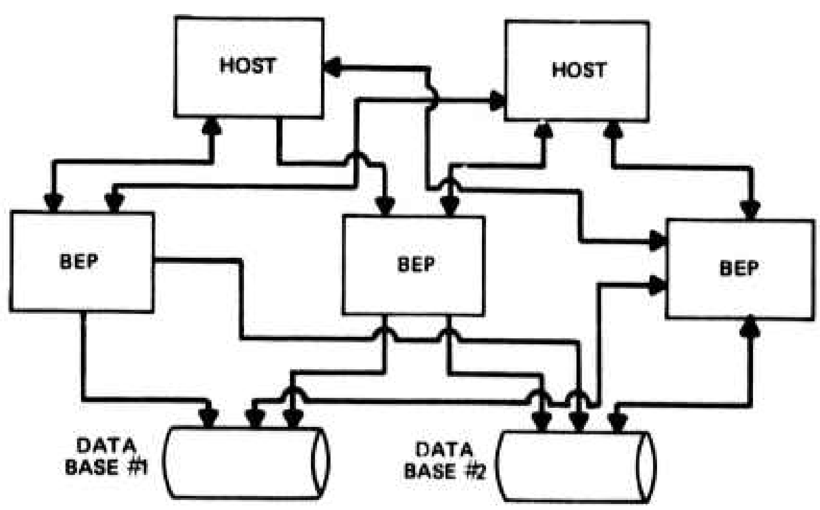


Figure 10b. Distributed Multiprocessing BEP

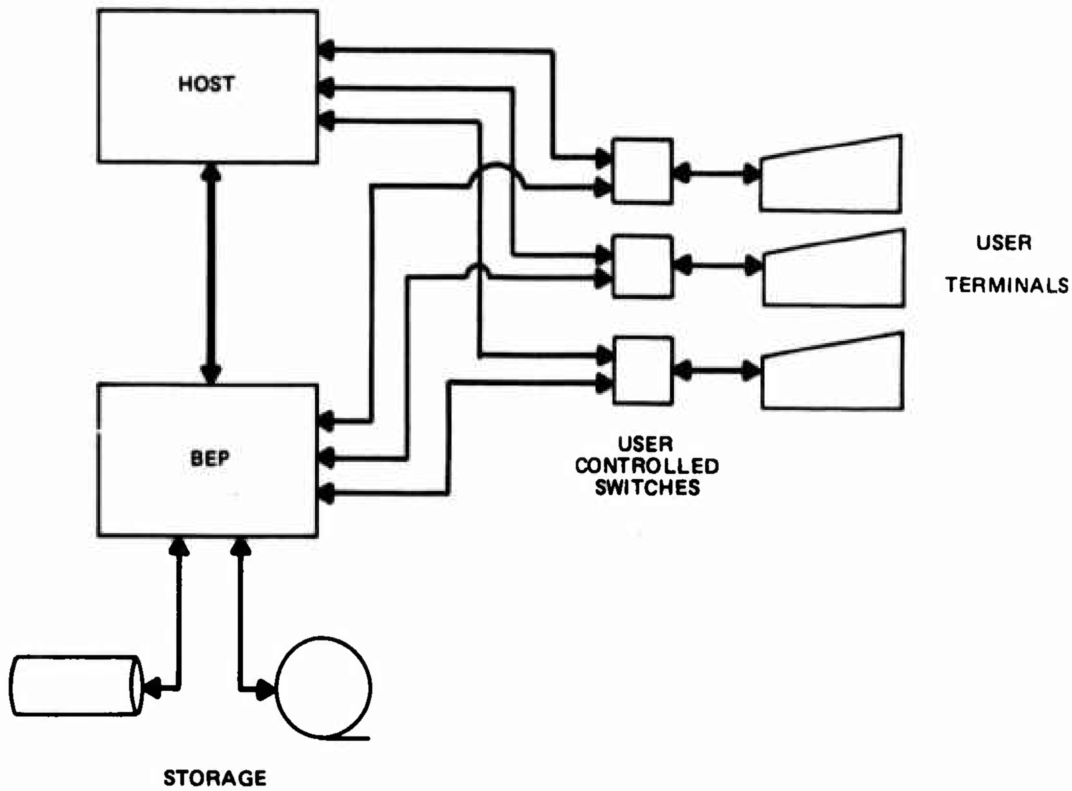


Figure 11. Partially Freestanding BEP

that has a lower cost per executed instruction than does the host. As mentioned earlier, the specific requirements for the BEP will depend upon the choice of DBMS, but a partial list of some desirable features might include: large word size (32 bits) to eliminate the need for memory management software; large maximum memory size (256K - 1M bytes); high I/O data rate (at least 4 Mbytes/sec); micorprogrammability; multiprogramming support; and of course, low price (preferably under $100,000 for a CPU with at least 256K core).

No matter how low its price, the BEP will still cost money. Furthermore, in most cases, the BEP and the host will be procured from different vendors; consequently, maintenance will cost more, and effective maintenance strategies may become more difficult to implement. Nevertheless, the BEP will be more cost-effective at data base management than the host, so relative savings should accrue.

Since the BEP will be responsible for controlling the storage devices on which the data base resides, it is essential that the BEP have the capability of interfacing with high performance, competitively priced disk and tape drives. If the system contains a mass storage subsystem (MSS) with large capacity (greater than 10^{11} bits) and slow access times (averaging more than one second to retrieve a file), then the BEP can also serve as a controller for the MSS and staged storage devices. DML requests from the host would be passed by the BEP to the MSS; the MSS would return an entire file to the BEP, which would transfer the data to high-speed disk storage. The host could then have either fast, direct, and random access to the data or indirect access through the BEP (Figure 12).

Sharing storage resources in a multihost environment has always been a difficult problem to solve, but the BEP concept can provide an answer. Because it has sufficient computing power, the BEP is capable of intelligently, and dynamically allocating storage space or entire devices to host subscribers. In addition, a MP/BEP which is distributed so that each device can be accessed by more than one BEP has very high reliability characteristics. If one BEP fails, its work load can be accepted by another. Thus while system performance may be degraded, it will not be destroyed.

The communication link between the host and BEP presents more hardware choices. If the link is a 9600 baud telephone line and the host and BEP look like terminals to each other, then implementation is easy, but data throughput is very low. Conversely, a channel-to-channel direct link has very high transfer rates, but it may be exceedingly difficult (or impossible) to implement such a link without severely distrupting the operating systems of the host and BEP. This type of decision cannot be made until system throughput requirements are well established; whenever possible, advantage should be taken of the BEP's power to reduce the amount of data which needs to be sent to the host.

Software. The primary software impact of the BEP is on the host operating system and application programs. Since all of the storage device I/O control routines, file management routines, and DBMS routines are executed by the BEP, the host OS has more CPU time available. Of course, this extra time

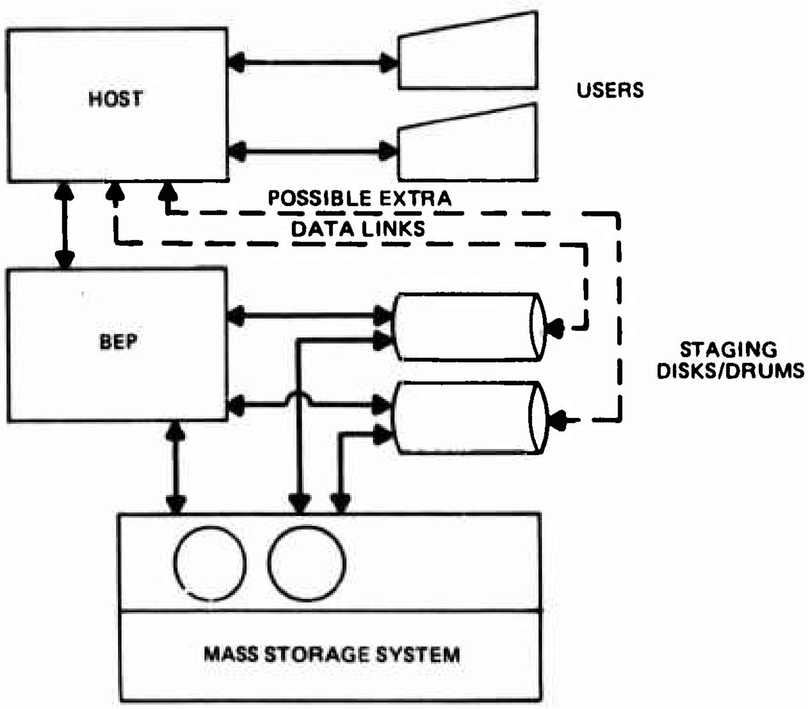


Figure 12. BEP as a Staged Storage/MSS Controller

is advantageous only if the host OS is multiprogramming, so that the host can continue to process other tasks while waiting for the BEP to respond to a DML request. Also, some of the extra time will be used by the BEP communications interface located in the host. However, the net resulting increase in host throughput and productivity may by itself outweigh the dollar cost of the BEP.

To users of the host who are accustomed to accessing the data base through a DBMS in the host, the BEP will appear to be totally transparent. If the host has had no DBMS prior to the acquisition of a BEP, then application programs will have to be modified and system users trained to make use of new procedures.

The BEP has its own complement of software of which the most important component is the DBMS. In fact, since the choice of a BEP may depend upon the desired DBMS, a particular DBMS should be chosen or designed as the first step in developing a BEP system. The factors involved in such a choice are outside the scope of this discussion, but it must be emphasized that the decision will affect the entire architecture and design of the BEP system.

Data Base. The existence of an integrated data base is implicit in the BEP concept. Without an integrated data base and a DBMS, the BEP is merely an expensive file control system. However, there is no need for the data base to be totally integrated; a MP/BEP can be designed such that each individual BEP manages a separate functional portion of the data base. While this approach is expensive, it may be a viable option, particularly when full integration of the data base itself would be costly or undesirable for other reasons (e.g., security).

A MP/BEP is also capable of supporting a distributed data base. For example, each of a group of remote BEP's could be dedicated to manage its own data base, and at the same time each would be communicating with a central BEP which would provide coordinating control for the entire BEP configuration. Alternatively, a large BEP could transmit data between a central MSS and remote staging areas. The exact type of data base distribution depends upon many management decisions which are too complex to cover here. Note that in any MP/BEP, especially one involving remote stations, time and space consuming network communication routines must be present in each BEP, thereby reducing the average processor efficiency somewhat.

Data integrity and security are major concerns of every data base administrator, and the BEP is capable of significantly enhancing system performance in these areas. The host and BEP can both maintain journal or audit files to keep track of data base changes. If either machine detects a failure in the other, then the system can be halted without compromising data integrity. The data base can be restored (once the fault has been corrected) by applying changes recorded in the audit files to an appropriate earlier backup copy of the data base. This technique will help avoid loss of data integrity due to system failure, since two cross-checking machines are better at detecting and correcting their errors than one possibly faulty self-diagnosing machine.

Data base security can be improved by the BEP in several ways. First, the only path to the data base available to users is through the host/BEP link. Even if a user gains control of the host's operating system, he would be unable to make illegal data accesses, since all data password and validation routines reside in the BEP, which is not under host control. Second, the BEP can constantly monitor the host for the occurrence of unauthorized operating system penetration. Of course, both of these safeguards are compromised if the BEP is either user programmable or partially freestanding. A third alternative is more expensive--provide a separate BEP and data base for each level of security. This procedure may be necessary when security regulations forbid the mixture of security levels in one machine regardless of how it is protected.

In addition to its direct impact on a system's hardware, software, and data base, BEP has far reaching implications for future system development. For example, the BEP can be upgraded without impact to host application programs. The BEP will enable system designers to take full advantage of the new storage technologies (e.g., CCD's, magnetic bubbles) by interfacing directly with the new devices. The host will continue to see a constant environment. An important consideration is the expressed willingness of the BEP vendor to interface his equipment with that of other vendors. Lack of such support from the vendor may result in a very inflexible development path for the purchaser.

4.1.4 Requirements for Back-End Processing. Given the great variety of possible BEP configurations, it seems intuitively obvious that some type of BEP architecture could be developed to improve the performance of almost any existing system. However, in order to maximize the benefits of BEP (and minimize its disadvantages), many different conditions should be met. While the relative importance of particular factors will vary from one system to another, the following list of conditions partially delimits some basic requirements for any BEP implementation.

- Requirements of the Host and Application
  - -- A DBMS or information system should exist and should be intensively used.
  - -- A major portion of the BEP total computer resources should be devoted to operating the DBMS or information system.
  - -- The application should be able to take advantage of the BEP's ability to reduce the amount of data processed by the host.
  - -- The BEP will have greater beneficial impact in an online, multi-user environment than in a single-user, batch environment.

- Requirements of the BEP Vendor
  - -- The vendor should be willing to adapt to changing customer needs and new technology (especially MSS and other storage technology).
  - -- The vendor should be willing to interface the BEP with several different types of hosts and peripherals.
  - -- Careful maintenance strategies must be implemented to ensure definitive fault isolation in a multiple vendor environment.

- Requirements of the Back-End System
  - -- The storage devices interfaced with the BEP should be competitive in both price and performance.
  - -- The host/BEP link must be reliable, and it should operate at the highest possible speed which would not require major operating system revision.
  - -- Most of the DBMS should be off-loaded to the BEP.
  - -- Application programs should not be off-loaded to the BEP.
  - -- The BEP should be responsible for data integrity and security.
  - -- The BEP should be multithreading, and when extremely high reliability or throughput is needed, it should also be multiprocessing.

4.1.5 Status of Back-End Processing. Since the concept of BEP has been developed only in the last four or five years, very few BEP systems have been implemented. Those which do exist are experimental or prototype models, and none of them are operating in a commercial environment. In 1972 Bell Telephone Laboratories, Inc., created what may have been the first BEP to be physically realized. The back-end was a Digital Scientific META-4 linked to a UNIVAC 1108 host, and the DBMS used was called XDMS, a modified version of UNIVAC's own DMS 1100. The purpose of this effort was to partially determine the feasibility of back-end data base management and validate its capability to improve total system performance. While the project succeeded in achieving these goals, it left untouched many important areas of study (e.g., MP/BEP or multihost environments, data security, and system reliability). One unexpected result was the demonstration that a workable BEP system could be established on a small machine (32K 16-bit words) in a relatively short time (18 months) with a relatively small manpower effort (6 man-years).

The only company to date which has announced an aggressive development program for BEP is the Cullinane Corporation. On the basis of an indepth evaluation of current minicomputer characteristics and capabilities, Cullinane concluded that the Interdata 8/32, the SEL 32/55, the PDP 11/70, and the MODCOMP IV/25 (in that order) are the four best candidates for BEP hardware. Although the Interdata 8/32 was found to be significantly superior to the PDP 11/70 (for BEP purposes), Cullinane has elected to focus its resources in the future on the development of its Integrated Data Management System (IDMS) on the PDP 11/70 as a BEP. There are powerful economic reasons for this decision: Cullinane had already invested heavily in IDMS on the PDP 11/70 as a freestanding system at the time they completed their minicomputer evaluation study; a large market for the PDP 11/70 already exists in both the government and commercial sectors of the economy. At the present time IDMS on the PDP 11/70 as a BEP prototype has been delivered to the DOD and is scheduled to be operational in April 1977. Commercial versions will probably be available in early 1978. The initial model will interface only with IBM systems and the development of interfaces with other hosts will depend upon future market demand.

There are indications that MRI Systems Corporation (a new corporation formed from elements of MRI and Precision Instruments) may also enter the BEP market. It is understood that the Interdata 8/32 has been chosen as the BEP to support MRI's System 2000 as a DBMS. Host/BEP interfacing software is expected to be

provided for most mainframes which now support System 2000 (IBM, CDC, UNIVAC). In addition, it is expected that MRI Systems Corporation will market the Precision Instruments System 190 mass storage device (section 4.2) along with staging disks as an integral part of its BEP system. This system would also mark the first marriage of a DBMS with a mass storage device.

In summary, BEP technology is still in its infancy, and as such its full potential will remain unknown for some time. However, any system engineer contemplating the future growth and complexity of data management operations on his system will be facing severe problems. Hopefully, the emergence of back-end processing as a mature technology will enable him to cope with those problems.

4.2 Mass Storage Systems. This section consists of a brief summary of existing mass storage systems and assessment of recent technical advances in this area. The summary consists of updated equipment descriptions which were originally included in RADC Technical Report RADC-TR-76-15, "System and Mass Storage Study for Defense Mapping Agency Aerospace Center (DMAAC)," January 1976. Extensive data contained in that report describing advanced technology and mass storage system implementation considerations has not significantly changed during the past year and will not be repeated in this report.

The DMATC currently has a projected requirement for storing as much as 4.7×10^{12} bits, by 1983. This volume of data, coupled with operational characteristics of projected drum, disk, and especially tape activities, was one of the primary motivations in entering into this mass storage study. It is readily apparent that some form of mass storage system will be needed to handle such a quantity of information. In this section we therefore discuss the developing technology which might provide such a system. Emphasis will be placed on the technology of particular mass storage devices in the environment of very large data bases (greater than 10^{11} bits). A brief assessment of future trends in mass storage technology is also included.

4.2.1 Ampex Terabit Memory (TBM) System

4.2.1.1 Characteristics. The TBM uses 2-inch wide, reel-to-reel magnetic tape as the basic storage medium. What Ampex refers to as a random access capability is provided by using tape search speeds of 1,000 ips (compared to 125 or 200 ips on conventional transports) and a packing density of 700,000 bits per square inch (compared to 14,000 bits/in^2 for standard computer tape). Data recording is done in the transverse mode (to the direction of tape motion) using a rotating head. This transverse scan rotating head and tape configuration is illustrated in Figure 13. As is evident, the relative head-to-tape speed is the sum of the longitudinal tape speed and the speed of the rotating head.

In addition to the transverse data tracks, the tape also contains three longitudinal channels: an address track used to identify individual data blocks; a control track needed for operation of the rotating head; and a tally track to record auxiliary information such as the number of accesses to a particular block, to identify the location of tape defects, or to identify tape sections which cannot be erased.

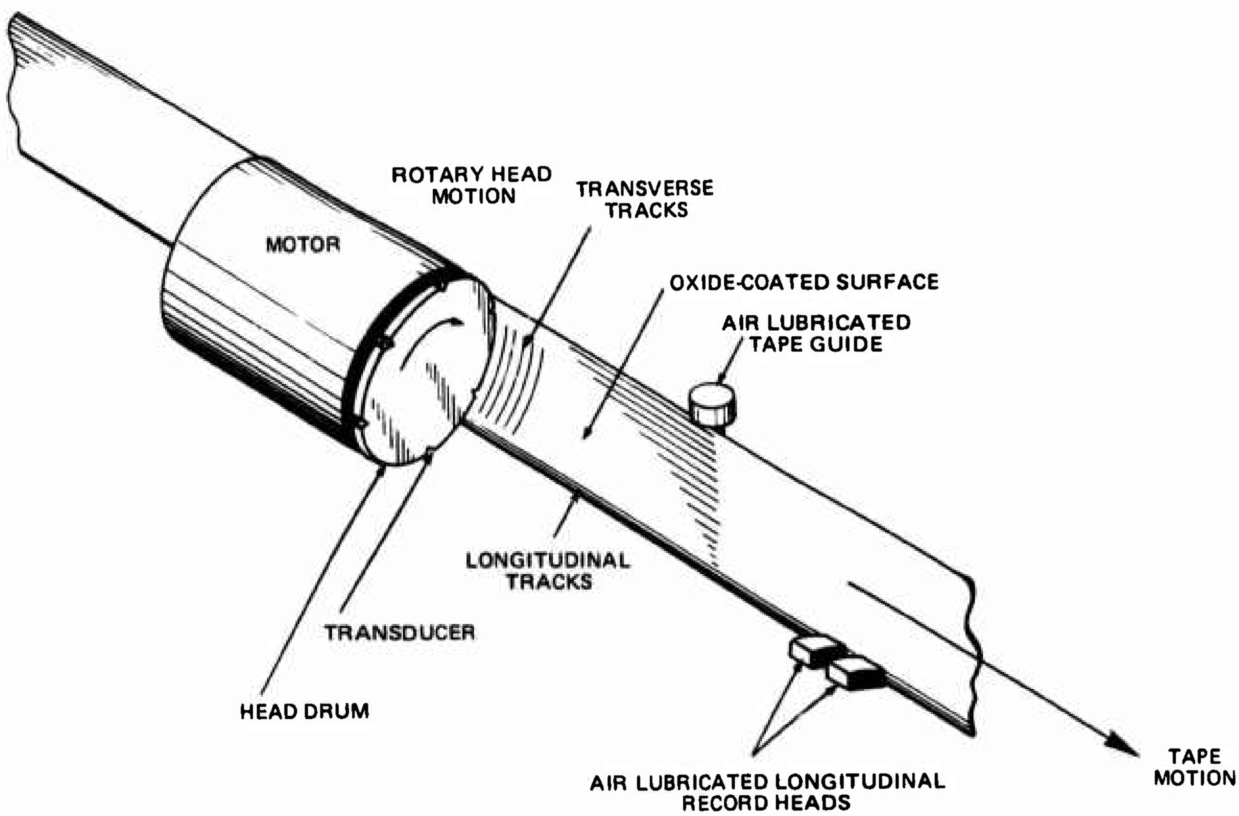


Figure 13. Transverse Scan Rotating Head and Tape

The standard tape configuration calls for a fixed data block length of 1 million bits. This translates into a length of 0.97 inches of the 2-inch wide tape. Each data block can be individually erased and recorded.

The actual memory system is composed of four major building blocks:

- The tape transport which contains the components necessary to move the tape.

- The transport controller which contains all of the circuitry needed to control the transport. Each controller can operate any one of the transports in the system.

- The TBM controller which is a general-purpose minicomputer provides command interface between the memory building blocks and the central processor unit as well as monitoring the internal functions of the TBM.

- The digital signal channel which accepts the digital input which is then FM modulated and written in that form on the tape. The read and write channels are independent thereby permitting simultaneous reading and writing operations on separate transports. Data erase is an independent function not requiring the data channels.

A typical TBM memory system consists of transport controllers, multiple tape transports, and data channels. In addition to the main components, buffers between the memory and the central processor unit are usually required. A switching matrix between the transport controllers and the transports is needed so that any controller may operate any transport. A similar switch between the transports and the data channels is also required.

4.2.1.2 Capacity and I/O Rates. The modular nature of the TBM permits the configuration of systems with a wide range of capacities and I/O rates. Each transport can accommodate up to 45,000 inches of tape, or about 5×10^{10} bits. Overall system capacity can be anything from 5×10^{10} bits to 3×10^{12} bits with 64 transports. According to Ampex the only limitation on system size is possible problems due to the placement of equipment and of the cables required to interface a very large configuration. For means of size comparison it should be noted that a reel of conventional 1/2-inch tape accommodates approximately 10^8 bits. Each data channel transmits at a rate of 6 megabits/second.

The following performance characteristics have been provided by Ampex for a typical 10^{12} bit system:

Number of transports	36
Tape per transport	30,000 inches
Number of transport controllers	5
Requests per hour	1200
Average time per request	3 seconds
Data transfer rate	6×10^6 bits/second

4.2.1.3 Reliability. The TBM system has been operational since 1972 and Ampex claims 98 percent uptime. The high degree of system modularity is probably responsible for this success.

Although hardware breakdown does not appear to be a problem, the mode of data storage itself may present difficulties. Videotape recording techniques (the source of the transverse rotating head concept) were never intended to provide the low error rates required for digital computer applications. To counter this, Ampex has employed a dual recording idea, i.e., each bit of information is recorded twice on the tape. The bit density of 700,000 bits/in^2 quoted earlier takes this fact into account and represents the usable density of information, the actual packing density being 1.4×10^6 bits/in^2. With this arrangement Ampex claims an error rate of 1 error in 10^{11} bits.

Another possible problem is that of tape wear. The rotating head is quite destructive to the tape and represents another imcompatibility between videotape and computer tape requirements. To at least partially counter this, only the control, tally, and address tracks are read during high-speed search. The rotating transverse head is not in contact with the tape during this opoperation. Ampex feels that this arrangement will permit at least 500,000 data passes to be made before reliability is adversely affected.

4.2.1.4 Software Impact. Ampex says that the command and control interface between the central processor and the memory controller is fairly simple, consisting of several simple commands such as "write address," "write digital data," "search for address," "abort last command," etc. However, Ampex literature cites the existence of interfaces only with IBM, CDC, and DEC equipment and it is therefore likely that the interface of a TBM with the UNIVAC mainframe would be a first-of-its-kind venture.

4.2.1.5 Cost and Status. A basic 88 billion bit TBM system can be purchased for $800,000; each additional 88 billion bits will cost $100,000. Current marketing of the TBM is being handled through an arrangement with the Systems Data Corporation (SDC). Under this arrangement SDC holds the marketing rights for Ampex TBM end-users, while Ampex will continue to sell the hardware system without software interface. This arrangement reflects a common problem facing all existing mass storage devices--the requirement for customized interface software. Under this SDC-Ampex arrangement a multiple TBM system is programmed for delivery to the National Weather Service, National Oceanic and Athmospheric Agency. Scheduled for delivery in June 1977, the systems will be used to create archival recordings of TIROS weather satellite images. Each tape record will maintain its own directory.

4.2.2 CALCOMP 7110 Automated Tape Library

4.2.2.1 Characteristics. The CALCOMP 7110 Automated Tape Library (ATL) is a fully automated online tape library for standard 1/2-inch magnetic tapes. Under computer control the ATL automatically brings the tapes from storage, mounts them on tape drives, dismounts the tapes when the job is completed, and returns them to storage. Accessing up to 150 reels per hour, the ATL can store up to 6,122 standard tape reels in a lockable self-contained library that can service up to 32 tape drives.

The system is modular. The basic configuration consists of one control unit, two storage units, a reel-selector mechanism and one automatic reel-mounting unit servicing one tape drive and storing 746 standard reels of magnetic tape. Sample configurations of differing capacities are as follows:

Control Unit	Storage Units	Tape Drives	Stored Standard Reels
1	2	4	746
1	4	8	1,514
1	6	12	2,282
1	8	16	3,050
1	10	20	3,818
1	12	24	4,586
1	14	28	5,354
1	16	32	6,122

These are not the only possible configurations. Use of thin cartridges can increase capacity up to approximately 9,000 reels.

The ATL control unit oversees all mechanical functions of the library and interfaces with up to 4 CPU's. The reel selector mechanism sweeps along at 100 inches per second, selects a tape reel from storage and places it in a premount station which automatically locks it on the hub of a tape drive. Once the tape reel is in the premount station, the reel selector mechanism is free to remove and return other tape reels in the library.

4.2.2.2 Capacity and I/O Rate. CALCOMP quotes a maximum system capacity of 9.6×10^{12} bits for the ATL. Maximum data transfer rate is given as 9.6 megabits/seconds. The access time for the mounting of a tape is given as 11.6 seconds for a system with 8 storage units and 13.6 seconds for a 16 storage unit system.

4.2.2.3 Reliability. There are 30 ATL systems currently installed in commercial and federal government installations. All of these installations are in IBM 360/370 environments. Commercial acceptance is viewed as significant. Users include the Atlantic Richfield Corporation and Mountain Bell Telephone.

4.2.2.4 Software Impact. The software for the ATL falls into two categories: hardware control and library management support. This software is designed to interface with IBM 360/370 OS and OS/VS. No interface with a UNIVAC 1108 system exists at this time, but CALCOMP has advised that one is currently under consideration by the Bureau of the Census.

4.2.2.5 Cost. The approximate cost of a 5,000 reel tape library is $300,000. Cost of a minimum size unit (746 reels) is approximately $200,000.

4.2.3 Control Data 38500. The Control Data 38500 is presently produced specifically for the IBM System 370. The following statements pertain only to the IBM compatible system although it is expected that later systems for other mainframes will possess similar capabilities.

4.2.3.1 Characteristics. The 38500 records all data on small magnetic tape cartridges which are stored in cells contained in a mass storage file. This file is a storage magazine in which the cartridges are filed according to an X-Y grid. Any specific cartridge is selected by a mechanical arm and moved from storage to an entry station in an average of 2.5 seconds. Then, after a 2-second move to a read/write station, the cartridge is opened and its tape unwound and drawn into two vacuum columns while remaining attached to the spool in the cartridge. The system contains 2-4 such read/write stations and the operations of selection, transfer, reading, writing, and storage can be overlapped to maintain a continuous flow of data between the mass storage facility and any disk of the central computer system.

The basic facility has a capacity of 2,052 cartridges. The effective recording density of the tape is 6,250 bpi and each cartridge contains 150 inches of tape resulting in a capacity of 64 million bits per cartridge. Data is recorded on 144 tracks across the width of the tape, with 18 tracks being covered by one position of the read/write head assembly so that the head moves 8 times to fully cover the total of 144 tracks. The facility therefore combines disk and reel-tape technology resulting in sequential operation with limited random access capability. The read/write head configuration and tape format is illustrated in Figure 14. Data can be accessed directly without searching the entire length of the tape and a so-called "user direct code" provides access to specific records within a cartridge.

Data access may be accomplished either through the use of a staging disk or may be read directly into central memory and returned directly to the mass-stored cartridge without the use of intermediate storage. Dedicated staging drives are not required.

Figure 15 is block diagram of the 38500 mass storage system. The mass storage facility contains the storage controls, mass storage adapters, and mass storage files. The mass storage adapter provides control and attachment facilities between the mass storage file and the System 370. Each mass storage file has its cartridge files, selector mechanism, and two to four read/write cartridge transports. Movement of the magnetic tape cartridge within the file is effected by the cartridge selector, and the read/write transports make data available to the System 370.

A CDC mass storage facility may be configured in almost any size, from a 128 billion bit capacity upward. It is possible to configure for the number of data paths, the number of read/write transports, the amount of redundancy approprite to the system, and the access performance required. Field expansion is also provided by module add-on.

The user's current direct-access subsystem may serve as the staging device for the mass storage file; no controller upgrade or modification is required.

4.2.3.2 Capacity and I/O Rates. The entire 38500 system provides 128 billion bits of online storage, about the same as 6,400 mag tapes. According to CDC a maximum of 5 seconds is required to retrieve a given cartridge and move it to a read/write station. Once threaded, tape movement can be continuous or incremental with a maximum transfer rate of 6.5 megabits/second.

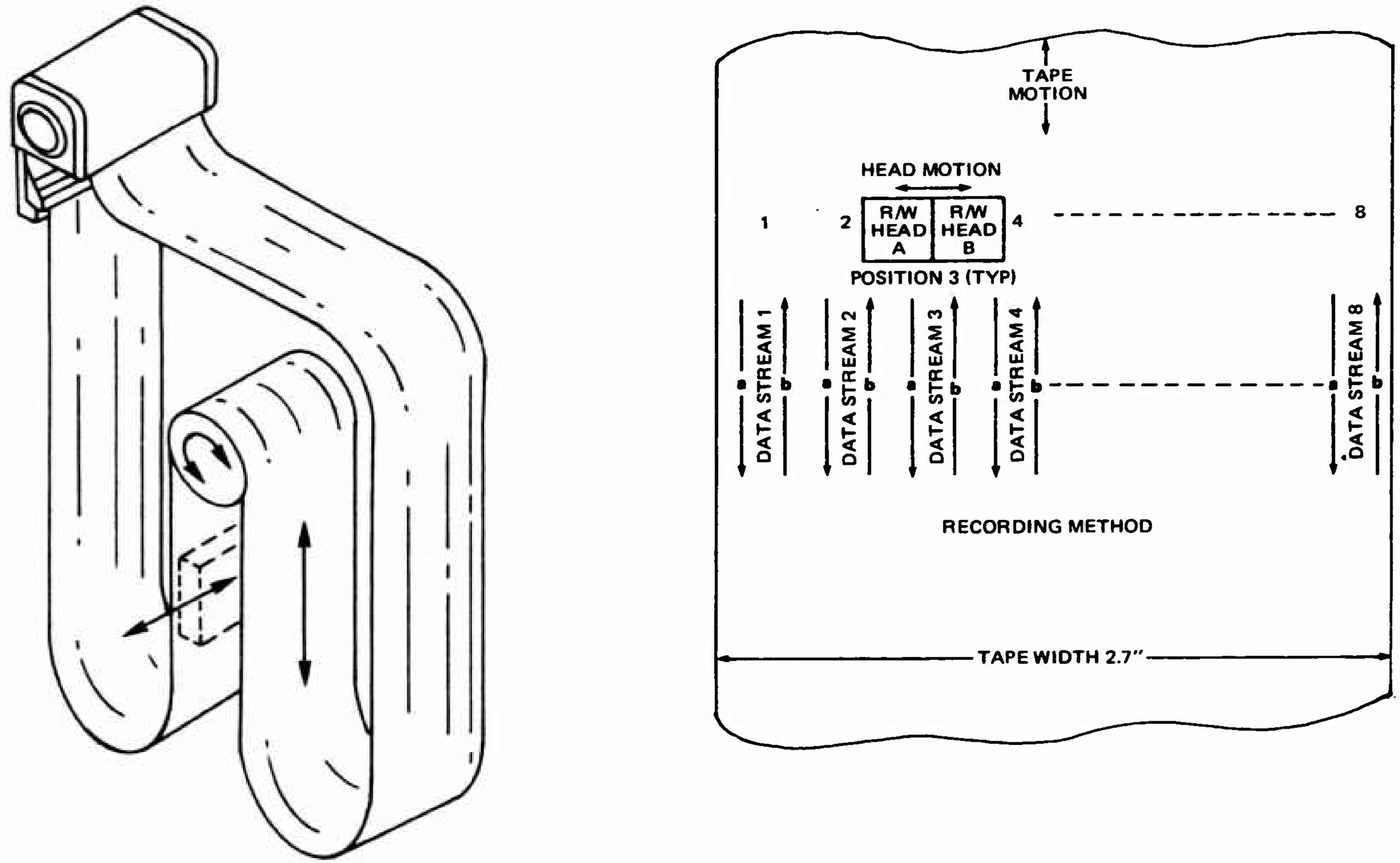


Figure 14. Read/Write Head Configuration and Tape Format for CDC 38500

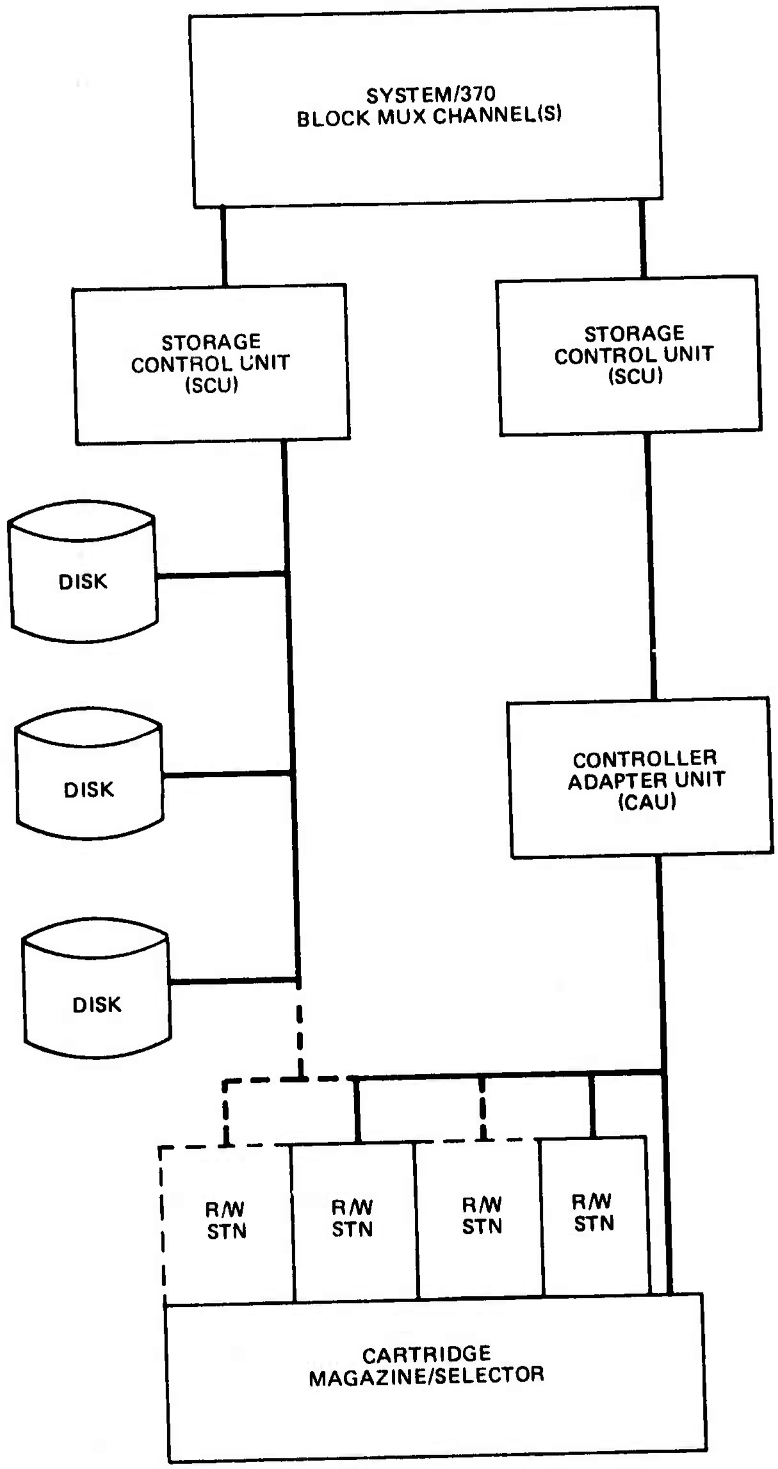


Figure 15. CDC 38500 Mass Storage System Block Diagram

4.2.3.3 Reliability. It is too early to make definite statements about 38500 reliability. However, the mechanical complexity of the device in regard to cartridge selection and tape read/write operations suggests that this unit may require careful maintenance.

4.2.3.4 Software Impact. Programming support is provided by the CDC Virtual Data Set Access Method (VDAM) software package under the control of OS/MFT, OS/MVT, or OS/VS software systems. VDAM controls the storage of data, remembers the location of data sets, and handles the staging of data on disk and the queuing of accesses. It will support user sequential access (where the user knows the location of the data sets) and user direct access (where the user calls for a block within a known data set but does not necessarily know the location of the set.)

The main point is that this system is presently offered only for IBM systems, later to be made in models for CYBER 170 mainframes. It may be some time before a UNIVAC compatible operating system is offered.

4.2.3.5 Cost. The basic 128 billion bit system with two tape read/write stations is priced at $326,335 plus $987/month for maintenance. Two additional drives may be added for $52,000 each. Cartridges are available in packages of eight for $120 per package

4.2.4 Grumman MASSTAPE

4.2.4.1 Characteristics. The MASSTAPE system employs data cartridges stored in a cylindrical array. Delivery to the read/write station is performed by rapidly rotating this cylindrical array and electromechanically positioning the desired cartridge into the read/write transport which is located atop the cartridge array. Each array contains 44 cartridges and 8 such arrays are physically packaged into one storage unit yielding a modular storage increment of 94.4 billion bits.

Each cartridge contains 260 feet of 1/2-inch tape, in a reel-to-reel configuration. Data is recorded at 8,000 bits per inch resulting in a real packing density of 256,000 bits per square inch. In normal operation there is 130 feet on each reel at the start of a file access and since the tape is passed at 150 ips, it takes a maximum of 10.4 seconds plus 1.5 seconds for delivery to the read/write station to read any bit of data in the system.

In order to determine access time to the start of a file one notes that each cartridge is subdivided into 288 data sections, 9 on each of 32 half-tracks. These data sections are storage allocation units of 10^6 bits each, allocated to one file at a time, and are analogous to cylinder or track allocations on a disk. Thirty-two data sections are available within 1.5 seconds and the average data section start point is obtained in 5.6 seconds.

4.2.4.2 Capacity and I/O Rate. Running the tape at 150 ips with a density of 8000 bits/inch fields a single file data rate of 12M bits/second. The MASSTAPE buffer unit accepts data from up to eight tape transports in whole allocation units of 10^6 bits. It is partitioned in such a manner that

16 simultaneously active files can be accommodated. The buffer device is presently a drum with a capacity of 17 x 10^6 bits together with five independent 8K x 16-bit core buffers. Grumman estimates a sustained throughput of 3,080 kilobits/second per buffer device.

Each storage unit of the MASSTAPE system provides a storage increment of 94.4 billion bits.

4.2.4.3 Reliability. Error correcting codes which can detect and correct multibit burst errors are implemented in the MASSTAPE system. During any transfer of data, 512 data word segments are operated on by a cyclic fire code which has the ability to correct all burst errors up to 160 bits in length. This results in an error rate of 1 in 10^9. An additional order of magnitude improvement in accuracy is attained with retries leading to an overall error rate of 1 in 10^{10}.

Failure of any MASSTAPE component is detected by MASSTAPE resident software and reported to the MASSTAPE control operator. The MASSTAPE system will continue to operate as long as one of each essential component is functioning.

4.2.4.4 Software Impact. The MASSTAPE system simulates a pool of conventional tape drives and requires no modifications to user programs. As with all mass storage devices, software interface to the central processor must be provided and at present Grumman's emphasis in this direction appears to be toward compatibility with IBM mainframes.

4.2.4.5 Cost and status. A 112 billion bit MASSTAPE system is priced at approximately $400,000, and will be sold only to OEM purchasers. This again reflects the problems associated with interfacing a mass storage device in a user environment.

4.2.5 IBM 3850

4.2.5.1 Characteristics. The IBM 3850 and the CDC 38500 are very similar in concept and, as is clear from the model numbers, CDC (whose unit is newer) seems eager to emphasize this fact.

Data in the 3850 is stored in cartridges arranged in a honeycomb-type library wall. The library accessing mechanism travels at a maximum speed of 250 cm/sec in the X or Y direction. At this speed the arm searches across 50 cartridges of 400 Mbits each in 1 second or 20000 megabits of stored data per second in each direction. The combined motion provides a 7-second average access time.

The cartridges are cylindrical in shape having a diameter of 4.75 cm and a length of 8.9 cm. The area recording density is 3.36 x 10^5 bits/in^2 and the tape length per cartridge of 770 inches results in a volumetric storage density of 2.48 megabits/cm^3. As has already been mentioned, each cartridge has a capacity of 400 megabits.

Information on the tape spool is organized into "cylinders." A cylinder is the smallest unit of data which can be moved and has the same storage capacity as a 19-track cylinder on a 3336 disk pack--about 250,000 characters. Each cylinder is recorded at a fixed location, and specific cylinders can be located by identifiers along the edge of the tape. Cartridge capacity is 202 cylinders which again translates into approximately 400 megabits per cartridge.

IBM uses the term "accessor" to refer to the cartridge selection mechanism. This accessor delivers the cartridge to a data recording device of which there may be as many as four in the system. IBM employs a helical-scan transport as the data recording device and in this sense it is similar to the Ampex Terabit System. This helical-scan head is illustrated in Figure 16. Such a device does not require the head positioning function or the complex interleaved head design that a parallel drive requires, thus simplifying transport design and (supposedly) reducing costs. Another advantage of the helical-scan configuration is that the instantaneous data rate is independent of tape velocity, and a high average data rate requires only a low longitudinal tape velocity.

4.2.5.2 Capacity and I/O Rates. The current 3850 provides capacity configurations from 280 billion to 3,776 billion bits. Although actual data transfer rates are not available, IBM quotes an average access time of 15 seconds as measured from the cylinder requests to completion of data transfer for a one-cylinder data set onto a 3330. It will be recalled that one cylinder is equivalent to 400,000 bits.

4.2.5.3 Reliability. The mechanical reliability of the 3850 is a question that only time can answer.

In order to provide high reliability in data transfer, error correcting codes and associated circuitry are used which can correct up to 256 of 1,664 bits on-the-fly. All data leaving the mass storage facility are corrected, if possible, or tagged. Facilities are also provided to correct errors occurring during destaging. The design allows correction of single error bursts of up to 11 bits per recorded data block. In addition, label information on the cartridges is duplicated to prevent loss of an entire cartridge due to tape defects in the label area.

4.2.5.4 Software Impact. Data flow in the 3850 system is controlled by a 3830 microprocessor. This data flow is illustrated in Figure 17. In addition to data control, the 3830 also manages the virtual-to-real address translation. These virtual addresses must be converted to real DASD drive addressed and specific cylinders.

The system also contains a second microprocessor used for mass storage control (MSC). Its functions include: placement and inventory of all cartridges; space allocation on staging drives; initiating all stage and destage operations to be done by the 3830 microprocessor; control of the movement of cartridges; and error recovery procedures such as alternate device and alternate path entry. As the single primary control for the 3850 subsystem, this

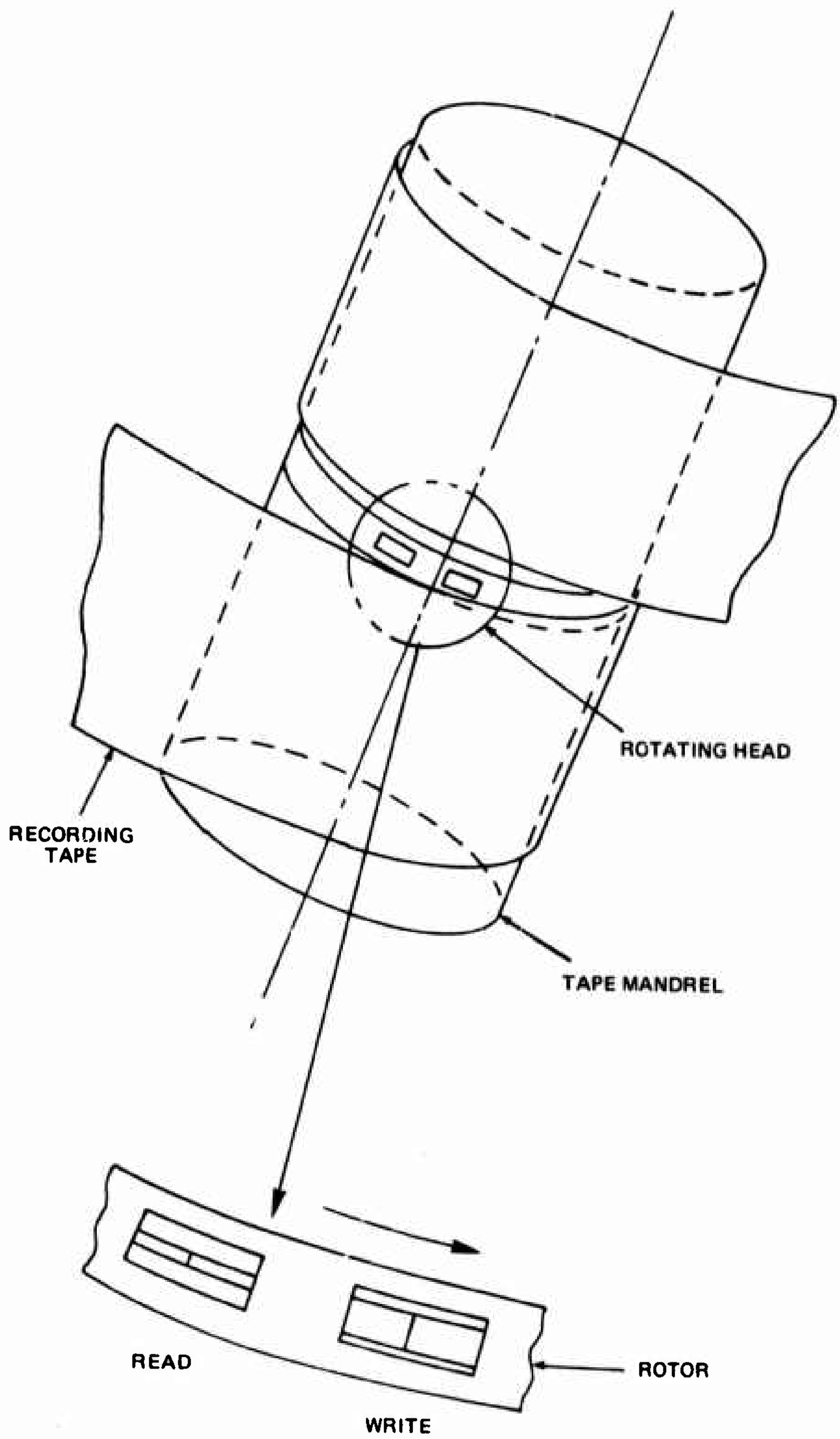


Figure 16. Helical-Scan Rotor With Head

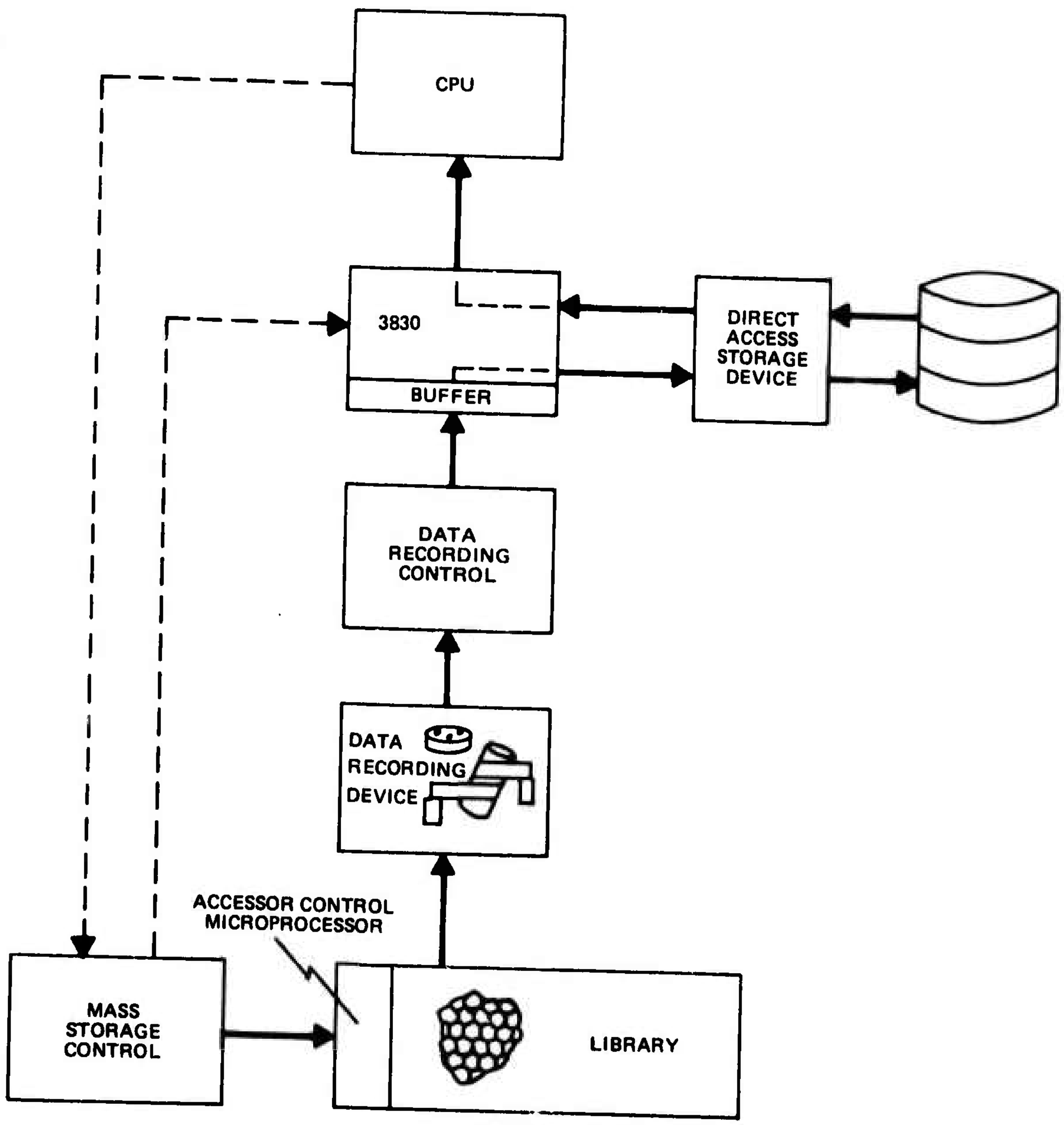


Figure 17. Data Flow Controlled by 3830 Microprocessor

microprocessor interfaces directly to the host computer, receiving commands for mounting volumes and acquiring data.

At the present time, the software interfaces required by the functions listed above exist only for the IBM System 370.

4.2.5.5 Cost. The basic component of the 3850 system is the 3851 mass storage facility whose purchase price will range from $477,000 for 280 billion bits to $2,304,000 for 3,776 billion bits. The 3851 contains the accessors, the accessor controls, and the mass storage control microprocessor. The 3830 storage control microprocessor is priced at $160,000. The data cartridges are $20 each.

4.2.6 Precision Instruments System 190

4.2.6.1 Characteristics. The System 190 uses a precisely focused laser beam to vaporize minute holes in the metallic surface of a special recording medium, the data strip, which is mounted on a rotating drum for read and write operations. By modulating the intensity of the laser beam focused on the rotating strip, a meaningful pattern can be generated in the media. During this writing process, light is reflected from the data strip, and the reflected beam is monitored in real-time to provide a read-while-write verification.

To read data, the incident light is reduced in power to a point where it no longer vaporizes the metallic surface. The light reflected from a previously recorded pattern is then monitored and used to reconstruct the data.

The basic recording media are optical quality polyester sheets, 31-1/4" by 4-3/4", called data strips. Onto each strip, a thin coating of rhodium metal is deposited and it is this coating which is selectively burned away in the process of data recording. Data strips, in quantities of one to ten, are mounted in a strip pack which provides a means of automatic handling and loading and serves as a clean environment for the strips when stored off-line.

4.2.6.2 Capacity and I/O Rate. Each data strip can hold up to 1.6×10^9 user data bits. Average access time has not been released but will probably be approximately 10 seconds for a 10^{12} bit memory.

For recording or reproducing sequential tracks of digital data, the maximum transfer rate of the laser recording unit is approximately four million bits per second, with an average error rate of one in 10^8 to 10^9 bits, depending on the data density selected.

4.2.6.3 Reliability. The method of data recording used in the System 190 yields permanent records which are not subject to degradation regardless of how many times they are read. In addition, the system has a simultaneous read-while-write capability that is unique to the laser hole-vaporization method and provides an almost instantaneous verification of data accuracy during the write operation.

Since the records are permanent, file updates on the 190 can be accomplished only by re-recording the entire data block on an unused portion of the data strip. Also, the effects of dust are critical in this system since a speck of dust falling into one of the vaporized holes can change the reflection properties during the read operation and thereby cause errors.

A recent announcement by Precision Instruments cites acceptance of a System 190 by a government agency on the basis of tests which "exceeded expectations in error rate, storage density, and transfer rate." Precision Instruments also announced that orders for a System 190 had been placed by the Energy Research and Development Administration (ERDA), the Heizer Corporation and Chase-Manhattan Capital Corporation. The latter two orders cited the need for permanence (non-erasable memory) in accounting records and indicated a total cost of $1.5 million. Some of these installations have since been delayed pending further enhancements of mechanical reliability. The division of Precision Instruments manufacturing the System 190 plans to merge with MRI Systems Corporation in mid-1977. MRI is a vendor of data base management systems and is best known for the System 2000 DBMS. MRI will provide software interfacing and market support for the System 190 as a component of a back-end data base processing system. This event marks an important milestone in mass storage systems--the increasing recognition that a terabit storage device is of limited value unless it is combined with a sophisticated data management system.

4.2.6.4 Software Impact. According to the manufacturer, the laser recording unit includes a programmable recorder control subsystem which can be designed to provide a hardware and software interface compatible with a specified computer system.

4.2.6.5 Cost. The cost of a basic System 190, including the system controller and a 128 billion bit read/write storage unit is approximately $400,000. However, it is unlikely that the System 190 will be marketed independently of its supporting data computer.

4.2.7 The Human Readable/Machine Readable (HRMR) System. The Harris Corporation HRMR system is presented separately in this study. While the HRMR is not a currently marketed system, it is in a late stage of engineering development by the Air Force and represents an emerging technology. The system has also been the subject of special interest for cartographic data storage. For these reasons, the HRMR system is presented here as a system of special interest.

The HRMR system represents a major development effort of the Rome Air Development Center towards achievement of a mass storage capability. Although the HRMR system originally addressed the document storage retrieval and dissemination problem and is primarily a read-only memory system, the permanent archival nature of the photographic recording media represents unique characteristics different from the wider range of currently available systems which utilize magnetic tape technology.

4.2.7.1 Research Prototype HRMR Microfilm Memory System. The research prototype of the Human Readable/Machine Readable (HRMR) Microfilm Memory System was developed by the Electronic System Division of Harris Corporation, formerly Radiation, for Rome Air Development Center. This was an initial step toward a mass memory microfilm data storage and retrieval system. It was delivered to Rome Air Development Center in May 1973 and has since undergone extensive testing and provides a vehicle for demonstration of the system capability.

The primary storage medium for the HRMR is standard 4" x 6" microfiche film. The equipment is capable of operating in three modes: an all digital or mass memory mode in which the fiche is used to store only digital data; a graphic arts mode in which both graphic arts pictures and the corresponding digitizer pictorial data are stored; and the alphanumeric mode in which the fiche is used to store HR alphanumeric data and the corresponding ASCII digital data.

The HRMR system is functionally composed of modules, each of which serves a special function within the total system. These basic modules are:

- Recorder
- Reader
- Controller
- Graphic Arts Scanner
- Microfiche Storage and Retrieval

Under a recent contract, the Rome Research Corporation performed a study and demonstration of the application of the HRMR device as a mass storage device for cartographic information. This analysis addressed the HRMR within the context of the Advanced Cartographic System (ACS) and included various aspects of input, output, application for product generation, and exploitation efficiency. The demonstration phase of the contract included the ability of the HRMR to process, record, recover, and exploit cartographic data for a variety of applications. Results of the study will be cited later in this section, and of necessity reflect only the capabilities of the first generation research prototype HRMR.

4.2.7.2 Description of HRMR Engineering Model Microfilm Mass Memory. In the Spring of 1974 the Harris Corporation began development of the HRMR Engineering Model Microfilm Mass Memory. This second generation system will be delivered in mid-1977.

The engineering model consists of a recording unit which formats, records, and verifies digital data stored on microfilm and a storage and retrieval unit which retrieves and reads the desired information. The emphasis of the program is to merge the previously developed HRMR techniques with large scale microfilm storage and retrieval mechanisms into a fully operational HRMR Microfilm Mass Memory. The HRMR Mass Memory System will consist of two major

subsystems: the HRMR Subsystem which will generate verified microfiche, and the Storage and Retrieval Subsystem, which will store the microfiche and provide rapid, random and automatic access to digital data recorded on the chips.

For this engineering model, the recording and readout rates, and the number of bits stored per microfiche is higher and the reliability and maintainability of the overall system should be improved. Technologies developed over the past several years are incorporated into the system to improve overall operational performance. Specifically, use is planned of an acousto-optic laser beam deflector in place of the multifaceted rotating mirror to scan the holograms onto the microfiche. Acousto-optic deflectors will permit higher recording rates while minimizing the maintenance associated with high-speed moving part deflectors.

Functionally, the engineering model HRMR Mass Memory System can be divided into two distinct subsystems. The HRMR Subsystem accepts data from any of several input sources, formats this data, records it on a film chip and verifies the recorded data. The Storage and Retrieval Subsystem stores film chips and, upon command, delivers a film chip to a data readout station.

A block diagram of the entire Mass Memory System is shown in Figure 18. Central to the HRMR Subsystem is the Controller Unit which consists of a Digital Equipment Corporation PDP 11/45 CPU, core storage, disk and tape storage, and I/O peripherals. The CRT terminal is a vital part of the Controller since the Controller, and, hence, the entire HRMR Mass Memory System, is interfaced to the operator by means of the display. All user input requests and system display outputs are also accomplished by means of the display.

A magnetic tape unit serves as the primary data input/output device. Data files, which are created externally on other user equipment, can be transferred to the HRMR Subsystem for processing and recording. Similarly, files which are stored within the HRMR system can be transferred to a magnetic tape for off-line utilization.

Another source of data input/output to the HRMR system is a direct user's computer communications channel. This channel can be utilized for file transfer in a manner similar to a mag tape file transfer and for facilitating direct online access of an existing user information data dissemination system to the HRMR Storage and Retrieval Subsystem.

For information which is not in digital format, such as line drawings, cartographics, or other black/white hardcopy material, a Digitizer Unit will be included as a part of the HRMR Subsystem. Digitizing of hardcopy source documents will be an off-line process to the HRMR Subsystem; therefore, the system Controller will not be required to execute in real time the control and data processing tasks associated with the digitization process.

Digitizer mag tape files will be transferred to the HRMR Subsystem for data reduction, processing, and fiche generation. Off-line digitization offers additional flexibility since multiple digitizers may be employed to increase

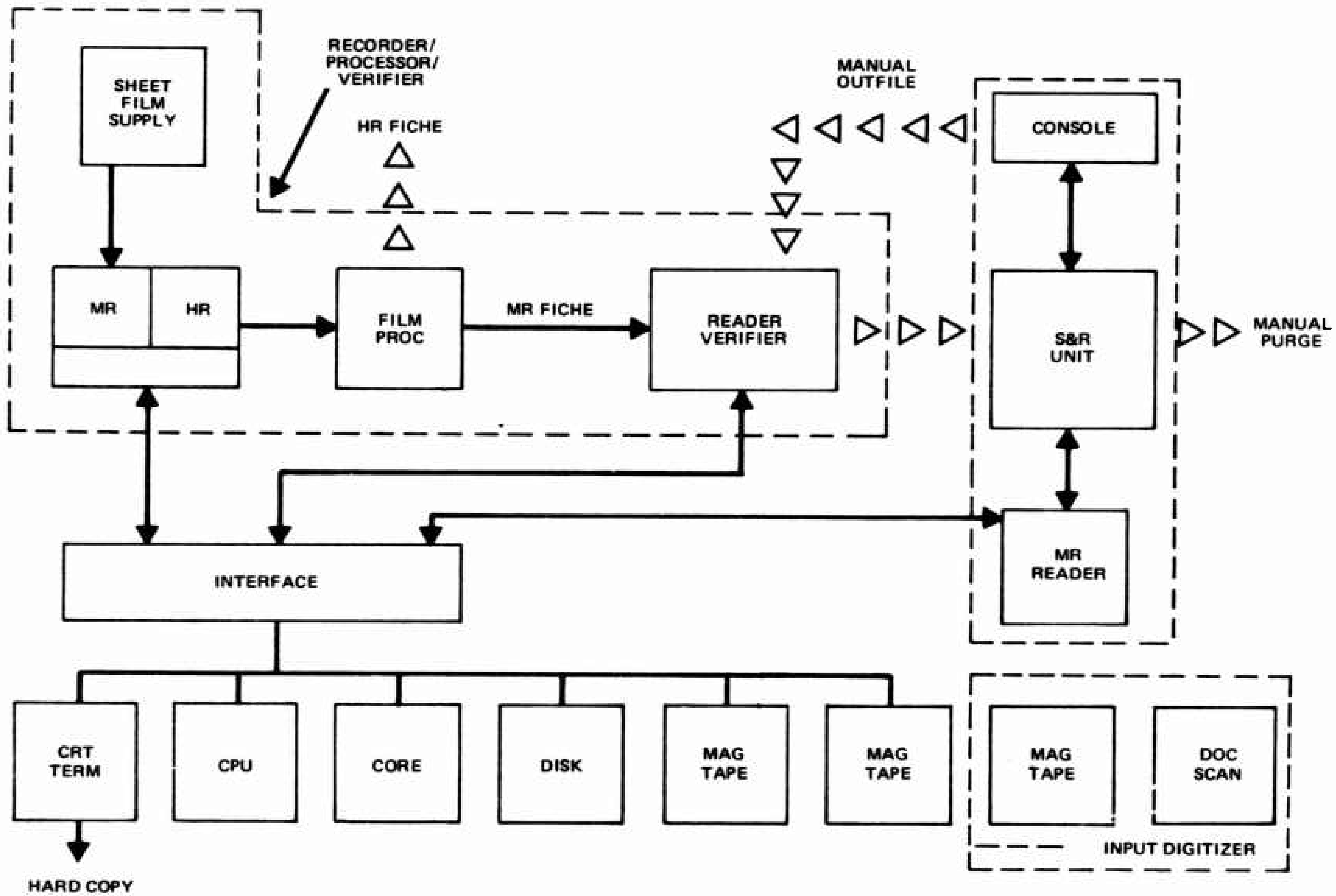


Figure 18. HRMR Engineering Model Microfilm Mass Memory Block Diagram

data input capacity. In addition, if the digitizer output is recorded on magnetic tape, it can be utilized directly on other user peripheral devices such as COM or display equipment.

The most critical part of the HRMR Subsystem is the recorder unit and its associated film chip on which all system information is stored. Film chip processing and digital data verification will be online, automatic, and an integral part of the recording process. To provide this additional capability, the engineering model recorder will consist of three distinct subassemblies grouped into a single physical equipment unit.

The first of these is the recording subassembly which will be basically the same as the present prototype recorder and will use similar film handling and recording techniques. Some modifications to the prototype model device implementation are being considered, especially in such areas as the laser scanner and modulator where increased performance is needed. Secondly, a film processor subassembly will be interfaced to the recorder and will provide online processing of the exposed film chip. Thirdly, a verifier station will be incorporated into the recorder unit. The verifier will automatically sequence through digital data files recorded on the fiche and will verify that data is properly recorded and processed. Verification of digital data is especially important since film supply imperfections or recorder performance deterioration can be quickly identified and corrective action taken. Since the verification task requires that digital data recorded on the fiche be read, the verifier can also function as a digital data reader. Other standalone reader units can be easily added to the system if future applications require such a configuration.

The most critical effort on the current contract is the integration of a holographic digital reader with a large-scale storage and retrieval system. Harris plans to interface the holographic reader with a modified fiche carousel system developed by Image Systems, Incorporated. This basic storage and retrieval system stores up to 750 microfiche in a single carousel. By stacking several carousels, an automated storage and retrieval system will be implemented which has the capacity of storing in excess of 10^{11} bits. Modular expansion of this basic approach may be a cost-effective way to achieve an intermediate-sized storage system.

Table 9 summarizes the performance goals of the HRMR engineering model Microfilm Mass Memory.

4.2.7.3 Summary of HRMR Cartographic Applications Study. Results of a study conducted by the Rome Research Corporation addressing potential application of the HRMR system for the storage and retrieval of cartographic data is contained in the final report to Contract F30602-75-C-0008, ACS Data Bank Development. The full report is relevant to the general scope of this study from a viewpoint of mass storage technology and potential applications of the HRMR system in the DMA environment. A review of the full report is recommended to interested personnel.

The following are extracts from the conclusions of the study:

Table 9. HRMR Engineering Model Microfilm Mass Memory Performance Goals and Parameters

Storage Medium:	105 mm x 148 mm microfiche
Storage Density:	30 megabits/fiche in MR mode (62 - 484K bit files) 60 alphanumeric images at 20X plus MR
Record/Read Rate:	500 kilobits/second within a file
Total Fiche Read Time:	2 minutes/fiche
Error Rate:	10^{-7}
S&R Subsystem Capacity:	6750 microfiche (2×10^{11} bit capacity)
Random Access Time to a Fiche:	$<$15 seconds

- "The hologram concept of digital data storage is an attractive means of achieving a high bit packing density while providing for minimal info degradation during handling."

- "The fact that machine readable data stored on microfiche is permanent (archival) gives fiche an advantage over magnetic storage devices which are susceptible to erasure and may require long term exercise or refresh. This permanency results from the fact that the data is stored on film, in a dispersed holographic form. In generating this film time must be allocated for photographic processing and parity checking. This additional operating requirement along with the inability to rewrite or change the machine readable code, may be a disadvantage for some applications, while other applications may not be affected by these procedures."

- "HRMR II is projected to demonstrate improvements over the original model, HRMR I, with improved recorder and reader hardware, storage and retrieval hardware, higher densities, and error correction codes.

- "From the cartographic applications surveyed at DMA, the use of human readable information does not appear to be necessary (from a data management point of view) on a microfiche that also contains machine readable data. From a data storage point of view, purely machine readable data is much more efficient."

- "There is a potential threat to inventory control and security in permitting someone to remove a microfiche containing valuable data from the storage and retrieval system in order to view the microimage under the present HRMR system. The risk lies in the possibility that the fiche might be lost or stolen or returned to the wrong location."

- "Possible applications of HRMR II or an improved version of HRMR II include the data bases at DMAAC (Defense Mapping Agency Aerospace Center) required by the Digital Land Mass Simulators and the Automatic Cartographic System." Note: Although not addressed in the conclusions, the report addresses potential use of the HRMR for storage of DTED and GIST data. The study also discusses potential means of providing source data to tactical field army users.

- "HRMR II could be a stand-alone system with data passed to and from it by magnetic tape. This, however, may be less efficient, due to the redundant data handling, (i.e., writing and reading of tape) than the alternative to let the HRMR II system, with its controlling computer, be a peripheral storage device to a major computer installation, such as the Univac 1108, with a high-speed data link."

- "If large amounts of data must be delivered from one installation to another, then, depending upon the cost of a microfiche machine code

reader at the destination, microfiche may be an attractive transfer medium. Also, if microfiche may be duplicated accurately without the involvement of a computer, as in photographic methods, then copies of the same data may be delivered to multiple destinations with no additional CPU cost."

- "There are only four other mass storage and retrieval systems actively on the market today; namely, IBM, CDC, Ampex, and California Computer. All of these are systems with very high total cost, limited speed and uncertain mechanical reliability. They are therefore of questionable value for 'on-line' operations, where demands of reliability and speed of access to data are beyond the capabilities of existing computer peripherals which have less mechanical complexity. Because of their high total cost and limited installation experience, they are not as yet economically proven for 'off-line' operation. The HRMR offers the potential for an economical cost savings medium for very large data bases in space and in unit price of the microfiche, and therefore might be competitively justified 'off-line'. Since the HRMR II is complex and unproven in mechanical and data reliability, it is initially considered restricted to near-future 'off-line' archival operations. The next phase, after its reliability is proven 'off-line', might logically be to interface to customized 'on-line' operations."

4.2.7.4 <u>Summary.</u>

- The HRMR system is an archival system with the unique capability of storing both digital and human readable (including graphic) information.

- A second generation HRMR system will be delivered to ADCOM for test and evaluation in mid-1977. This system will provide a more firm basis for the demonstration of system performance and mechanical reliability factors.

- Based on the study performed by the Rome Research Corporation it appears that additional modifications of the HRMR would be desirable for use of the system to store cartographic data.

- Cost of the HRMR system remains difficult to establish. Current best estimates place the probable cost of production versions in the $750,000 range.

- Based on current delivery and test schedules it is estimated that production versions of the HRMR system could be available around 1980.

- Based on the current assessment of DMATC mass storage requirements, which are predominantly of an archival nature, continued monitoring of the HRMR system for potential use as a mass storage device is warranted. The absence of any optimal solution to cartographic

data storage problems will require an open attitude towards all mass storage technology.

- As a low cost, permanent storage media the microfiche could provide a means of digital cartographic data for field army exploitation. While an evolution towards this type of application would take considerable time, an open attitude towards its use in such a manner should be maintained.

4.2.8 Future Mass Storage Technology. During the next decade researchers in storage technology will achieve orders of magnitude of improvement in the storage characteristics of most media. The two important aspects of all such technologies from the perspective of mass storage (greater than 10^{11} bit) systems are their cost per bit (including the cost of all mechanical/electrical/optical hardware needed to access and move data or control its placement) and their volatility (the rate at which the decay of the storage ability of media creates errors in the stored data). Volatility tends to be a fixed property of a given type of storage media, whereas cost per bit can be decreased by increasing the density of data stored per unit area. There are theoretical limits to the data density achievable in any recording surface, but no technology has reached its limit yet.

4.2.8.1 Magnetic Tape. This is a mature and highly competitive technology which will probably continue to be the most cost-effective choice for mass storage systems for at least the next five years. Current costs are as low as 10^{-4} cents/bit, and they will drop further as both linear and areal bit densities increase. The new 6,250-bpi tape drives currently on the market represent such a step in the development of standard half-inch tape, and 10,000- to 20,000-bpi tapes are likely to become available in the early 1980's. Because mechanical placement tolerances will decrease, the very high density tape systems will employ sophisticated EDAC routines to ensure low error rates. Videotape based systems (like the Ampex TBM) will also be improved during this period of time. Consequently, any potentially competitive technology will probably have to achieve 10^{-5} cents/bit in quantities of greater than 10^{12} bits in order to match the cost of future tape systems. Since this prospect is unlikely, magnetic tapes will continue to be the primary choice for large mass storage systems.

4.2.8.2 Disks (Magnetic and Magneto-Optic) and Drums. Magnetic disks and drums are a well-proven technology for storing large volumes of data with relatively quick access times (10 to 50 milliseconds). However, they are not yet practical as online devices for terabits of data. For example, about 400 spindles and packs of the new, high-density IBM 3350 disks would be needed to store 10^{12} bits of data online. This figure is clearly unacceptable for reasons of cost and size. If the density of such disks could be increased by another order of magnitude (an event which might easily occur within 5 years), then the hardware requirements for storing 10^{12} bits of data would fall to 40 spindles and packs. Thus it is very possible that magnetic disks will be used for terabit data storage in 5 years, and tapes will be reserved for extremely large (10^{14} bit) data bases.

Magneto-optic disks are still in the research stage, and it is unlikely that they will develop into a reliable device for mass storage systems during the period of time covered by this study.

4.2.8.3 Optical. Although optical data storage has the potential for higher bit densities than does magnetic surface recording, no reliable digital optical storage systems exist today. A practical erasable storage medium for true holographic memories has not yet been developed; and although Holofile Industries, Ltd., has announced the creation of a microfiche-based non-erasable holographic storage device, it has not yet released the stage of development of the device. Synthetic holographic systems (e.g., the Harris HRMR system) and non-holographic systems (e.g., the PI 190) both suffer reliability and error rate problems as the result of high mechanical complexity. While any of these problems might be overcome very soon, it will take years for the technology to demonstrate its commercial viability and safety.

4.2.8.4 Solid State Storage. Many types of solid state devices, including MOS semiconductors, CCD's, magnetic bubbles, etc., have been proposed as possible storage components for large volumes of data. However, none of these devices will achieve a low enough cost per bit for trillions of bits to compete with tapes or even disks. The main comparative advantage of solid state devices versus mechanical access devices is the rapid access time, a characteristic which is relatively unimportant in mass storage systems.

4.3 Security Issues. In this section we recognize the issues surrounding the problem DMATC faces protecting and securing a portion of its data and computer programs. A discussion centering around the types of security, the locations of greatest concern, and alternative security measures follows.

The basic conflict pits security consideration against efficient usage of computer resources. Because only a portion of the DMATC operation is considered classified, the question arises: How can DMATC be certain classified information will not be purposely or inadvertently accessed and compromised without causing an inefficient computer configuration? As DMATC presently has their configuration, two completely separate installations provide the security assurance. There is no question, given a non-hostile environment, that dedicating one machine to classified operations is a secure arrangement. However, this situation causes a duplication of certain resources without the expanded capability the addition of these resources could provide if they were intertied and sharable. Therefore, this discussion is based on the assumptions that:

- The computer system and the programs with data will be stored and used in a non-hostile environment

- The program/data mix involves both classified and unclassified information

- Only one computer system is used with many sharable resources (e.g., network)

- The system is multiprogrammable, possibly a multiprocessor

Given the above assumptions, there are four primary locations where security measures are of paramount importance: main memory, secondary storage, transmission channels, and terminals. With both classifications of information possibly coexistent in main memory, certainly some protection mechanism (e.g., segmentation, security kernel) is necessary. Likewise, it is certainly reasonable to assign secondary storage media individual security classifications. In the case of magnetic tapes, cartridges and removable disk packs, a dichotomy of classification need not introduce much inefficiency. However, it will still be necessary to provide protection (e.g., capability matching) from inadvertent secondary access when the secured media is online. Some measure is also required to ensure the classified data is clear from memory and scratch disks. Transmission channels, on the other hand, do not require much attention in a benign environment. Generally speaking, appropriate encasement of such cables or the use of fiber optics should suffice. And, finally, at the user's terminal the user and terminal may be categorized for secure operations via various types of protocols (e.g., password).

A well established conceptual model for security control information is a three-dimensional access matrix (Figure 19). On one axis of the matrix are subjects (e.g., users); on another axis are the objects (e.g., program files, data files), and on the third axis are the capabilities (e.g., read, write, password). Entries in the matrix are Boolean values, indicating whether a capability is available to a subject for a given object. For most systems the matrix is rather sparsely populated, with subjects having access to relatively few objects. Therefore, actual implementation will use some more compact and logically equivalent data structure.

One of the alternative approaches specified for consideration in this study is a transition to a distributed processing capability. A distributed system is one in which the control of data storage and processing is decentralized and information must be transmitted between system segments. A multi-user, multiresource, multisystem environment needs more than simple physical and procedural security mechanisms. Some alternative security techniques are presented in the following paragraphs to support a position that security technology does exist and research continues to improve and reduce the costs of these techniques.

Recently (in "Electronics", Vol 49, September 30, 1976), an Air Force-Mitre Corporation team has announced the development and verification of a "security kernel" for a Digital Equipment Corporation PDP 11/45 minicomputer. The kernel is a hardware and software device which ensures that users can get in and out of multi-security-level data base only the information authorized at their levels of security clearance. The kernel requires less than 1,000 program statements and about 17,000 bytes of storage. The developers are working toward specifications for similar security kernels which will work in a variety of commercial computers.

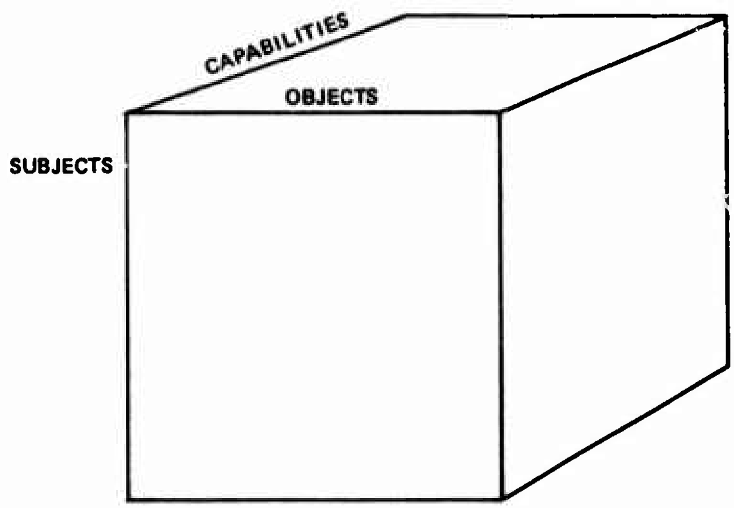


Figure 19. Capabilities Access Matrix

Perhaps a more relevant development involves an approach to computer network security using cryptographic devices. The work done by Heinrich and Kaufman combines a special purpose processor, the network security center (NSC), and data protection devices called network cryptographic devices as pictured in Figure 20. The network front-end is a processor which implements connection-oriented functions for a set of terminals and hosts. The network security center centralizes the authorization process. With the assistance of the cryptographic device (i.e., master) and stored access control information (e.g., capabilities), the NSC authenticates and permits connections between network resources. There are also cryptographic devices (i.e., slaves) at each network resource.

The cryptographic device encrypts data (i.e., transforms data to a cipher) and decrypts (i.e., reverses the encryption). These transformations are based on a secret parameter called a key. The master cryptographic device coupled with the NSC manages the establishment of new keys at the slave devices. If a slave device is attached to a single terminal, it may maintain only one key. However, if attached to a host or a network front-end, a slave device must be able to maintain several new keys to distinguish between various logical connections of that resource.

The basic encryption device used in the system described above makes use of the National Bureau of Standards Data Encryption Algorithm (this Federal Information Processing Standard was approved November 23, 1976). The characteristics which make this standard well suited for network security include:

- The secrecy of the transformation is dependent only on the secrecy of the key, not on the secrecy of the algorithm.

- Key length is 64 bits (8 are reserved for parity). There are 2^{56} potential keys. The length of the key is not amenable to exhaustive search techniques, and yet transmission to remote devices is not hampered.

- It does not require position or time synchronization and is independent of communication subsystem. The algorithm is block-oriented (i.e., 64-bit blocks).

- Single bit changes cause unpredictable effects on the output text. Enciphered text pairs do not aid code-breaking. Penetrators are forced to use exhaustive techniques to break the code.

- The algorithm is available in an LSI package making the device inexpensive to purchase and easy to implement and add to a network configuration.

Security issues become more impressive the closer the architectural configuration approaches a network configuration. However, in a benign environment with some combination of capability modeling, kernel technology, cryptographic devices, and further technology advancements, the concern for the

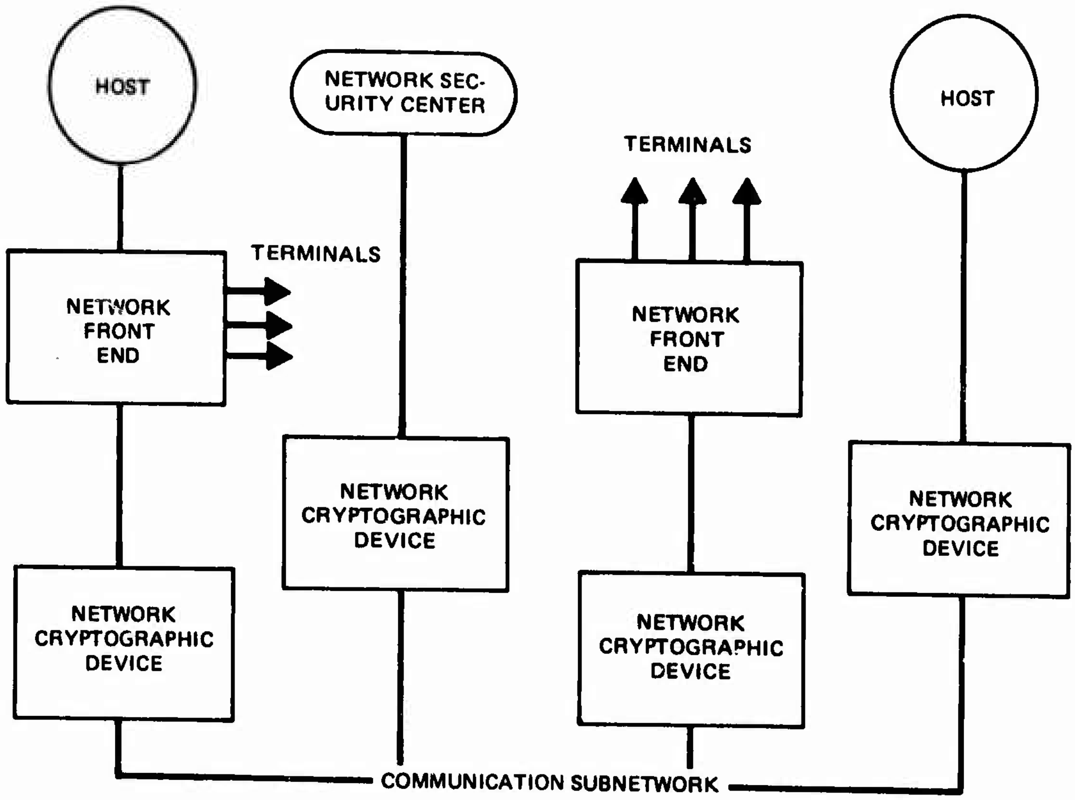


Figure 20. Network Security

possible compromise of information is lessened. With the possibility of SAA data being accessed and transferred to memory within any computer, the requirement for advanced hardware/software security technology packages greatly increases. Certification of a system that has access from terminals outside the SAA-cleared area requires that there be extensive testing. A result could be that all computers and terminals must be located in an SAA environment, and data leaving this area must be manually "sanitized" by authorized personnel. Utilizing past experience, present expertise, and future technology, a system can be designed that incorporates the required techniques to provide the system security necessary for the DMATC environment.

This discussion of security is recognized as brief and only highlights the issues involved. We contend that technology now holds a realistic potential for providing a solution to the control of multi-level security information. This problem is considered the greatest single constraint to the adaptation of advances in distributed processing technology in the DMATC environment. The same constraint would also exist if a centrally hosted time share environment were to be pursued and can be expected to constitute a constraint during the establishment of an integrated information management structure. Development, testing, and acceptance of suitable techniques capable of supporting a distributed processing capability will require extensive lead time. Pursuit of a viable solution to security issues should be made a priority DMATC objective.

5. ARCHITECTURAL ALTERNATIVES

This chapter reviews the major considerations which most significantly affect an advanced system architecture, discusses possible alternatives in this area, and concludes with a detailed description of an architectual approach which we feel will support DMATC requirements.

5.1 Factors Influencing System Architecture. The preceding sections have addressed a broad scope of activities having varying levels of influence on development of an advanced system architecture concept. These factors include the existing production environment, projected requirements for processing and data growth, promising areas of advanced technology, and long-term objectives of the Center. Functionally these factors represent computer processing needs, a need for more response information management, improved means of storing and retrieving growing amounts of data, and changing methods of the cartographic production process itself. To further complicate the issue, a recognition of lead time requirements can lead to conclusions which are somewhat contradictory.

Within this broad scope of influencing factors the following considerations were considered the most significant in terms of defining a system architecture which would support an orderly growth throughout the time frame of the study.

5.1.1 Computer Work Load Requirements. The rapid growth of processing requirements projected by DMATC will require substantial augmentation of central processing capabilities.

- The rapid rate of growth will require system augmentation during FY 77.
- Growth, in both the SAA and collateral areas, will require that capabilities be augmented in each area.
- SAA processing requirements will require an augmented capability approximately twice as great as the existing UNIVAC 1108 capacity operated on a 3-shift, 5-day week basis.
- Collateral processing requirements will approach the estimated UNIVAC 1108 system maximum capacity in FY 77 and exceed that capacity in FY 78.

5.1.2 Need for Work Load Distribution. A conclusion reached by the study team is that improvement in production throughput times will require a transition toward distributed satellite processing systems to perform data edit and formatting functions in an interactive mode. Specific areas in which this type of system should be considered are related to cartographic data generation (SACARTS) and photogrammetic post-processing. While this conclusion appears contradictory to augmented central processor capabilities, lead times associated with the implementation of distributed processing subsystems will preclude immediate relief to the rapidly growing DMATC processing work load.

- Distribution of functions related to data capture is motivated by the extended calendar times associated with batch mode cycling through a central processor.

- Implementation of distributed satellite processing systems requires extensive lead time (two to three years) and offers no immediate relief for the rapidly growing data processing work load.

- When fully operational processing demands on the central processors should decrease. The extent of the decrease will depend on the functions incorporated in the distributed satellite processing systems. The most dramatic effect, however, will occur in production throughput in terms of calendar time.

5.1.3 Growth of Data Holdings. The projected growth of data holdings will require accommodations for efficient means of mass data storage. While an initial assessment of potential data holdings has been made, a continuous effort to verify this project in terms of quantity and response requirements is considered essential.

Projections for online data storage requirements by 1983 indicate a minimum requirement of approximately 225 million words, exclusive of catalogue file and scratch space. Data storage at this level is well within disk storage capacity.

Projections for tape holdings consisting of source data base files (UNAMACE masters, SACARTS, DTIB) could reach a level of 4.7×10^{12} bits by 1983. Within the time constraints of the study, this data is considered primarily archival in nature and selection of storage media will be primarily an economic consideration. Further definition of access requirements both in frequency and response time could significantly alter the selection of storage media.

5.1.4 Improved Information Management. The need is recognized for a comprehensive program addressing information management.

- Through a detailed analysis of files, existing independent files should be examined for currency, redundancy, and accessibility.

- Files should be selectively identified for:
  - Consolidation
  - Development of automated interfaces
  - Restructuring under a Data Base Management System

- System growth should include provisions to allow integration of a Data Base Management System.

5.1.5 Automated Chart Production. From a functional point of view the work load associated with automated production of the 1:50,000 topographic line map is a dominant factor both in terms of data processing work load and growth of data file holdings. (Attention is drawn to the Addendum to this report which cites a recent change in DMATC program direction.)

5.1.6 Security. Protection of multi-level classified information is and will remain a major constraint and consideration during the evolution of any DMATC system architecture.

5.1.7 Associative Array Processing. While the associative array processor has demonstrated a capbility to perform cartographic applications processing, the cost effectiveness of this approach has not been established. Conversely, the potential utility of this processing technology, particularly in advanced concepts of digital imagery processing, warrants its consideration as an option for the future.

An extended period of research and development must be anticipated before implementation in a production environment can be considered a viable option. Since associative array processing technology is still in the process of evolving, particularly from an applications point of view, a system concept which would allow but not require its future integration with the DMATC environment should be considered.

5.1.8 Objective of Pilot Digital Operations. The post 1985 objectives of the DMA as expressed in the PDOP also provide a set of growth guidelines toward which system development in the 1977-1983 time frame should be directed. Of particular interest are objectives which are to be pursued concurrent with but outside of efforts specifically directed toward digital imagery processing. Included within these activities are the following.

Integrated Data Bases. Achievement of these objectives will result in a large volume of information transfers. This activity can be reduced by combining redundant data into integrated files. Nonetheless, there will be:

- Significant transitory file handling, mainly involving highly volatile (temporary) data
- Increased online storage of data requiring high-frequency access
- Archival (low-frequency access) data storage with appreciable growth expected
- A considerable file design and software development effort to improve methods of controlling and accessing data elements (data base management)

Distributed Access. Such PDOP subsystems as data base query, update, and purge or preproduction image evaluation--which assume distributed access to data bases--and preparation of press plates or direct printing from digital data--which assume distributed processing--imply that DMA's data processing operations be tied together by some kind of network.

5.2 Central Processor Augmentation. Prior to addressing a long-term system architecture concept, the rapid growth of data processing requirements forecast by the Center requires immediate attention. While this course of

action addresses only one aspect of the requirements scenario, its satisfaction is basic to maintenance of a capability to support production efforts while longer term solutions evolve.

5.2.1 Projected Work Load. The central processing capability must be able to support a work load which will double during the current year with potential for tripling during FY 78. The requirement for processing during 1977 exceeds the 1976 work load by a minimum of 100 percent. This requirement is viewed as primary and basic to DMATC mission accomplishment.

The need for a strong central processing capability remains an initial architectural consideration both in terms of the capacity required and the lack of time for development and implementation of viable processing alternatives.

Satisfaction of the FY 78 central processing work load will require an increased CPU capacity together with significant improvement of I/O capabilities. As the work load increases on augmented central processors, a corresponding need for increaded I/O capacity will have to be addressed.

5.2.2 Central Processor. The optimal approach to meet projected processing requirements within the required time is an augmentation with a UNIVAC 1100/40 class processor. This approach provides:

- A modularly expandable system
- Maximum compatibility with existing software
- Upward compatibility with current UNIVAC 1108 peripheral devices
- State-of-the-art capability with maximum life expectancy

The UNIVAC 1100/40 and UNIVAC 1110 both represent an approximately 100 percent increase in computer speed over the UNIVAC 1108. A detailed trade-off analysis will be needed to select a specific system configuration and can be initiated on the basis of benchmark testing which has already been undertaken. Since the greatest growth is associated with SAA production requirements, use of the augmented capability for this purpose would indicate optimal configuration with a 2 x 1 configuration. Augmentation of collateral processing resources through coupling of the existing UNIVAC 1108 processors should adequately satisfy validated requirements at this security level and leave undisturbed existing external communications channels.

5.2.3 Increased Information File Transits. Along with the increase in DMATC production demands, supported by improved CPU capacity, one can expect a proliferation of tape activity. This will be accompanied by an increase in the level of usage as well as the quantity of the dedicated I/O systems at the Center (digitizers, plotters, etc.). The focal point for these movements of files from system to system will remain the central processor complex. Current and future plans call for expanding the role of minicomputers and dedicated mini-based processing stations with equipment such as the DIODE and improved cartographic digitizing systems. At present all interfacing

between these devices and the large central host computers is by tape, which involves a somewhat error-prone human intervention component when high levels of activity are involved. The extent to which tape handling activities can be expected to increase can be readily appreciated by consideration of current activities related to plotter operations. This activity also illustrates the significant level of data preparation and formatting activity associated with a graphic function in the cartographic production process. Work load for the major cartographic plotters typically supports some 2,000 plots per month. Preparation of drive tapes on the UNIVAC 1108 and the physical movement and control of these tapes between the computer facility and the centralized plotter facility represent a significant part of the tape related work load which will grow in direct proportion to map sheet production.

Proliferation in tape activity with the augmented configuration can be expected to emphasize the apparent contrary tendencies of increased centralized and distributed environments. Our attention is then forced to focus on the common thrust in these two areas: the transiting of information files between equipments. The optimum accomplishment of these transfer operations then becomes a key requirement in architectural renovation considerations. This requirement is best satisfied by minimizing tape activities and automating the information transfer. Given the additional requirements for increased storage capacity and data management capabilities, the staging concept presented in a following section appears to be an optimal solution.

5.2.4 Security Constraints. The processing restrictions associated with SAA data further complicate the advanced system architecture. The physical segregation of Site I and Site II has minimized the system time lost by shutting down and purging since only one system is affected. Growth of SAA processing and data storage requirements remain on a par with collateral activities and necessitate augmented capabilities at both security levels.

At least on an interim basis retention of an isolated central system operating in a dedicated mode appears to provide the best solution to this problem. Reversion to a system architecture which would require total system shutdown (i.e., only one mainframe) will become more costly in terms of lost production with an augmentation to an advanced higher speed processor.

As discussed in section 4 advances in technology are expected to provide viable means by which an integrated system may be developed without system shutdown for processing multi-level security data. With this technology it is expected that it will be feasible to integrate the two sites and ensure that security is met without a shutdown-purge mode of operation. Through a combination of software automatic purge and hardware technology including removable disk packs it will be feasible to restrict unauthorized access to protected data and clear the classified data from core storage.

Development of an acceptable means of providing multi-level security protection in an integrated enviornment is critical to the optimal utilization of advanced data processing technology in the DMATC environment. This is equally true for achievement of the objectives of the PDOP. It is our position that advanced technology will make such protective measures a reality within a reasonable time frame.

5.3 Proposed System Architecture. The preceding section addressed an augmentation to the existing DMATC central processing facilities. This course of action is recognized as a necessary expedient to provide a capability to respond to a rapidly growing processing work load. This approach by itself will not provide solutions to other Center objectives such as improved information management or reduced product throughput time. As was previously pointed out, augmentation of the central systems can be expected to significantly increase the volume of tape handling since this mode of operation will remain the primary means of information file transit. System augmentation similarly does not address the potential use of an integrated Data Base Management System, whether hosted on a mainframe or as a Back-End Data Base Management System: the use of associative array processors; or the use of mass storage systems.

To provide a system architecture concept which meets the overall objectives of the DMATC through 1983 and beyond, a modular form of distributed processing is considered most amenable to future growth. This concept will allow implementation on an orderly incremental basis while preserving flexibility and allowing integration of additional systems or components as specific technologies mature. The concept proposed is a simplified adaptation of system architectures being addressed in two highly complex environments. The first of these is the system architectural design developed for the Air Force Global Weather Central. The second is the Strategic Air Command intelligence data handling enviroment. Advanced network technology currently being developed in these environments will provide a solid technology base for pursuit of the system concept proposed for DMATC consideration.

The data processing hardware configuration recommended below involves two major innovative concepts:

- A computer intertie--a general electrical, protocol, speed, data and device interface normalization technique for a wide array of subscriber interfacing requirements. (A subscriber is any piece of hardware--satellite computer, terminal, host computer, COPE 1600, associative processor, CELESTE, plotter, etc.--attached to the intertie bus via a microprocessor port.)

- The staging computer--a computer and disk device which is shared, and thus accessible by both the intertied processors and the mass storage device.

Proposal of these concepts is the result of a preliminary overview and needs more detailed examination before final judgement on specific components should be made. Similarly, use of the term "intertie" is intended to describe a generalized function rather than any specific existing device.

The general structure of the proposed DMATC architecture is shown in Figure 21. We feel it is the type of system that fosters the distributed processing concept while still retaining the extensive capacity of large centralized mainframes. The focal point of this figure is the staging center to which all transit files are passed. This and the system-wide communication bus allow

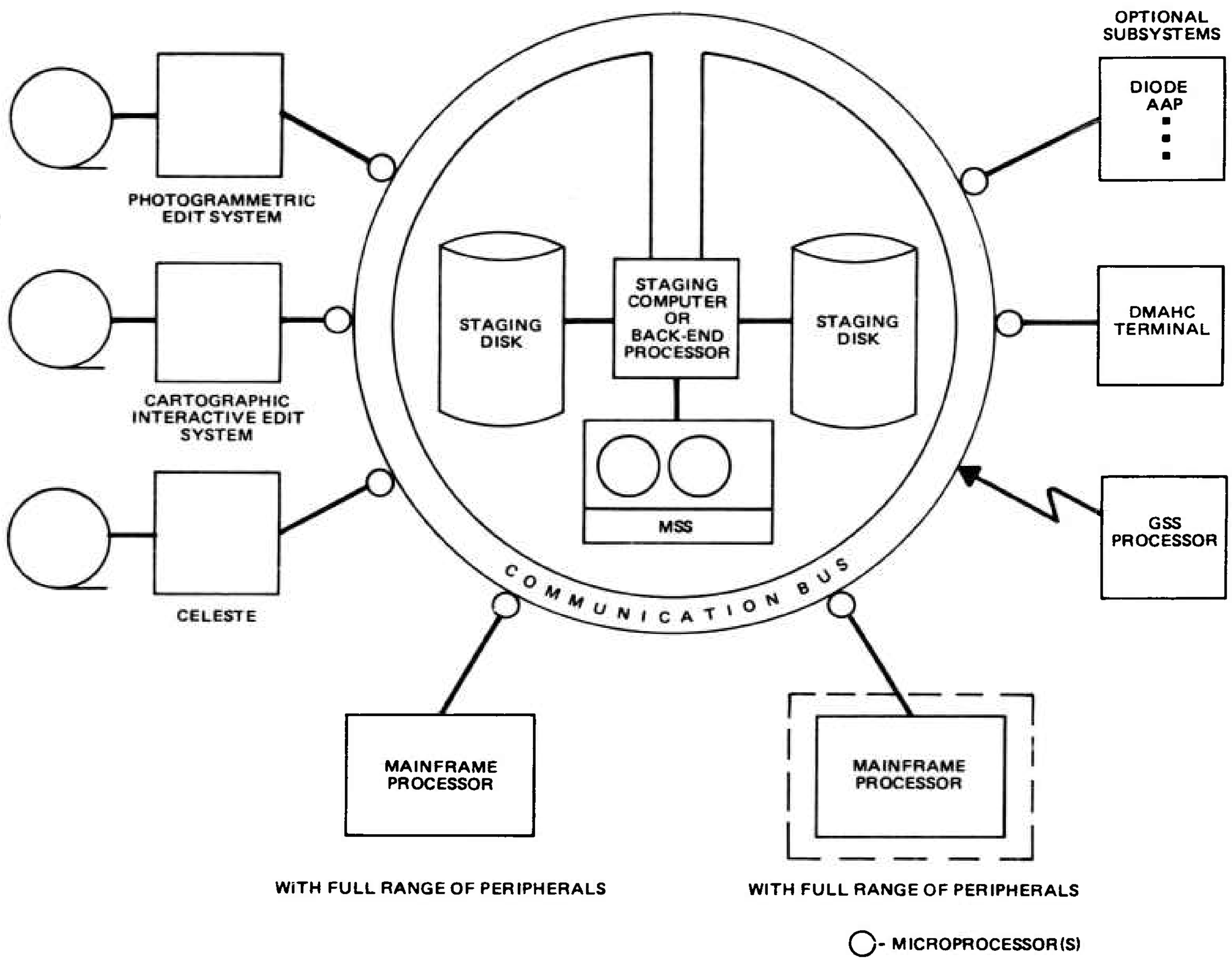


Figure 21. Proposed Advanced Configuration

all components in the environment to be tied together. We call this distributed processing through a computer intertie concept. It is supported by the semi-centralized general purpose mainframes, although relative to bus activies, all members attached to the bus appear equal. Much of the material which follows should be studied relative to this figure.

5.3.1 Justification and Assumptions. The overall goal of this concept is to provide a fast flexible data communications media between the diverse DMATC resources. The concepts presented here are designed to provide the optimal mixture of isolated modularity and interrelated communications and processing. The following points summarize the basic assumptions used to gauge the scope of this requirement:

- The amount and variety of data from dedicated I/O subsystems will increase substantially.
- Dissemination of finished cartographic products will increase in volume and variety.
- The amount of time required to produce finished products should be reduced to allow more sophisticated resource utilization.
- More raw computer capacity will be required to satisy increasing loads.

Providing a vehicle to meet the distributed system objective requires a design constrained by the following issues:

- Systems Security. Any intertie design must not compromise any security rules demanded by DMATC requirements.
- System Reliability. A distributed approach must rely on a stable communications capability if its overall mission is to be accomplished.
- Systems Performance. The intertie network should not introduce negative performance characteristics into the ADP operation. Data transfers on the bus should not adversely affect computer throughput.

The purpose of this section then is:

- To provide background of sufficient depth for effective decision-making on the alternative architecture
- To draw attention to the long lead time factors of equipment procurement which require timely decisions
- To illustrate that the bus concept is the basic common denominator for this architecture and should be given appropriate priority

The bus/intertie concept is implied by a broader requirement which in ADP terminology translates to making separate interdependent mission-oriented systems (the dedicated I/O systems) perform their collective function in support of the overall mission in more optimal fashion. Any scheme for responding to this requirement, of necessity, takes on certain ubiquitous connotations, i.e., the intertie media will be a constantly encountered factor in all information acquisition, information processing, information analysis, and information dissemination operations. For this reason the performance, reliability levels, and security protection supported by the selected intertie media must be of a very high order.

Effective steps toward problem solution necessitate certain assumptions about the available tools to work with. The following is an attempt to make some categorical statements about the ADP tools available for the intertie solution:

- ADP hardware in the 1980's will not be totally reliable. Any system design formulated must make provisions for the loss of any component without loss of system function.

- Sealing off errors and failures is within the capabilities of existing technology.

- Since electronics technology is imperfect, and undoubtedly will remain so, reliability can be achieved through hardware redundancy, which is consistent with the intertie design characteristics.

The above assumptions constitute an important aspect of this concept. High reliability operation of the intertie media must be a primary objective during design and integration. Considerable emphasis must also be given to error detection and correction procedures to minimize and isolate the impact of failures.

5.3.2 The Intertie Concept. The function of the intertie system is to provide the automatic intercommunication of diverse ADP resources within the DMATC environment. All data interchanges between intertie subscribers will be through the intertie media.

The intertie as conceptualized herein is made up of the following components:

- Bus. A bus is the means by which computer hardware devices are connected together, so that data may be transmitted from one to another. A bus might be made from wires, cables, or fiber optics light pipes. The type of bus being proposed for DMATC's future processing needs (asynchronous, bidirectional, bit-parallel, demand-driven, multiple-access) is capable of transmitting large quantities of information simultaneously; and the information may be enroute between many different and independent origins and destinations.

- Port. A port is a microprocessor subsystem through which a subscriber is connected to the bus. If specified, the port supplies the source and destination addressing attributes that messages

possess while on the bus and performs whatever conversions are necessary to ensure protocol, speed, and electrical compatibility between different intertied devices. Each port is a self-contained microcomputer system complete with CPU, bus, ROM, RAM and I/O interfaces. Port microprocessor busses are interfaced to the systems I/O bus via a bus interface module which controls the individual port's bus access. Traffic is multiplexed on to the system I/O bus to or from the staging area. (See Figure 22.)

- Bus Controller. Somewhere in the system a processor would need to be designated which would have overall control of bus activities and which would resolve and prioritize conflicts arising from simultaneous requests for use of the bus or ports. This responsibility could possibly be vested in the staging computer.

The intertie concept will make possible the following generic ADP capabilities:

- Network Storage Management for Distributed Data Storage and Retrieval. Through the central staging computer, any authorized subscriber attached to the bus will be able to address the data originated by any other subscriber also interfaced to the information bus.

- Ability to Interchange Loading Assignments Between Hosts. The job queues and program files can be treated as additional files at the staging computer. The loading level at each host can then be "balanced" out so new jobs can be directed to the least busy host.

- Load Leveling Among Compatible Distributed Dedicated Processors. As peaks shift between information acquisition, information processing, and information dissemination, it will be possible via the bus to dynamically align resources and requirements.

- Ability to Predict System Saturation Levels. By offloading communications functions from individual hosts to microprocessors, compute-to-I/O ratios tend to be more stable, presenting a more streamlined loading picture for the host processors.

- Centralized Data Management. The control potential of using the Staging Computer as a Back-End Processor for data base management isolates this function. Imposition of file standards as well as localized data monitoring is therefore facilitated.

NOTE: The identification of a back-end processing capability in Figure 21 has been included to emphasize the need for centrally managed and controlled data base management. It should be noted that the proposed intertie concept is not restricted to use of a back-end processor to perform the data management function. Use of a mainframe processor as host for a Data Base Management System, though conceptually less desirable, would also be a viable alternative.

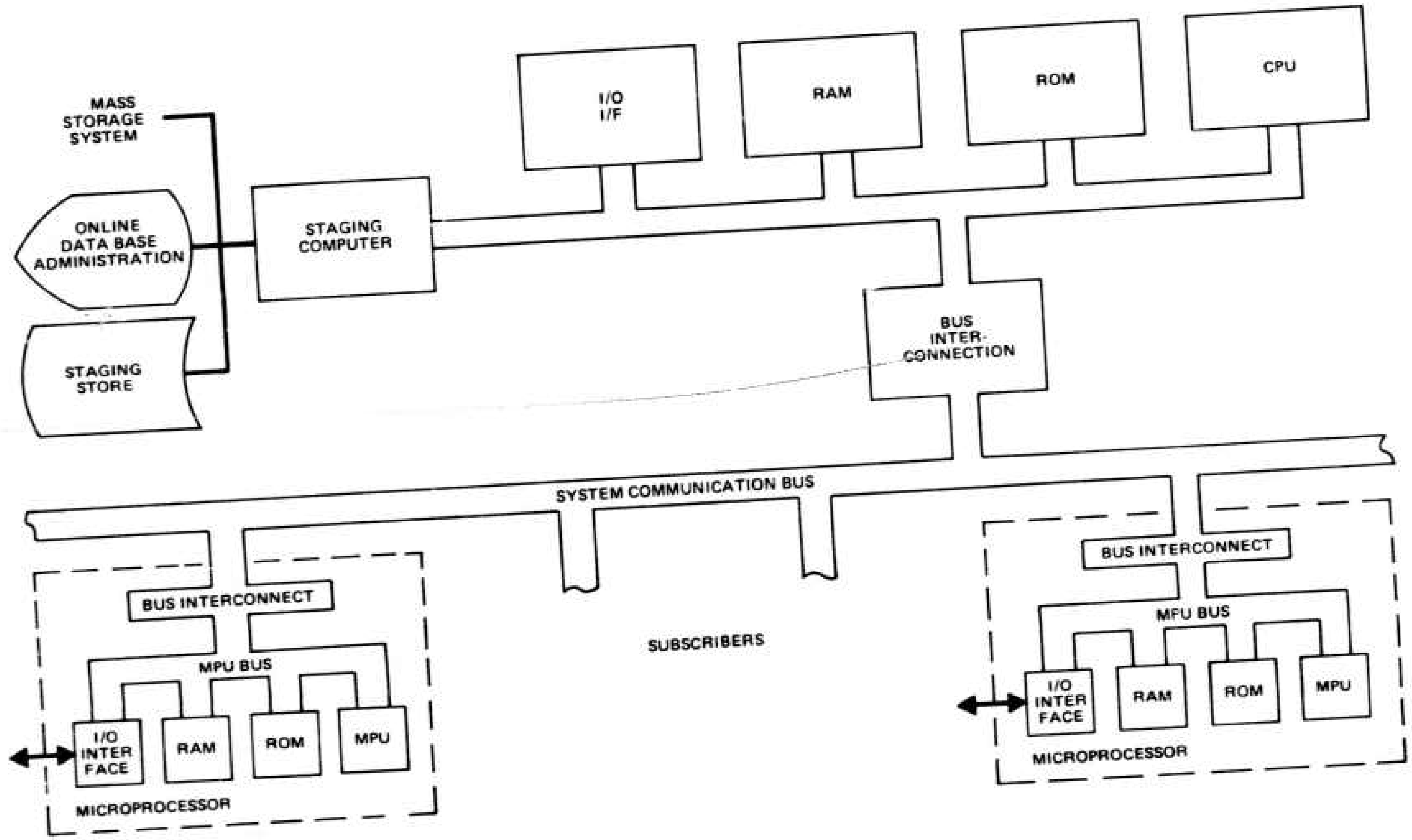


Figure 22. Microprocessor Unit Functional Flow

- Modularity and Expandability. Logical extension of the underlying architecture should produce amplified capabilities and adaptability for the future.

- Dynamic Configurability. If a system remains in an operational state 24 hours a day for perhaps several months, several abnormal capabilities are required, e.g., it may be necessary to attach or detach subscribers without interrupting intertie service. Thus, the attached subscriber entities must be able, and are required, to acquire their distributed configuration information dynamically during system operation.

- Multiplexing/Normalization. The bus provides normalization of speed, electrical, and protocol mismatches between transmitter and recipient.

5.3.3 The Staging Computer. This is the name given to a processor dedicated to managing file manipulations for, and file transfers between, mass storage and host or satellite processors. Files or data sets are transferred to tape or disk devices which are accessible to both the processors and the Mass Storage System (MSS), prior to or during program execution. The MSS, upon command, fetches the required files and transfers them to the mutually accessible disk or tape devices in a format ready for direct processing with no further MSS action required during execution. New or changed data sets are copied back into the MSS upon completion of program execution.

The staging computer provides almost complete transparency to the user for the following reasons:

- File request messages are transmitted to the MSS via the intertie.

- No host or satellite computer resources are consumed during the data transfer operation.

- There is very limited direct electrical interdependence between the staging computer complex and other processors. Thus, the MSS data connection is by way of the intertie just as though the staging computer complex were another subscriber in a multi-subscriber environment.

- The data is transferred to a standard, staging-computer-supported device so that no application or system source code modifications would be required for program execution; users would rely on standard and conventional access methods to process data.

- Minor external job control language changes might be required to indicate that certain files are to be MSS resident; but no device characteristics, addresses, or parameters would be required other than a file name (unless multiple mass storage systems were employed).

- The staging concept has the advantages of enforcing data base standards and of isolating any data base management system. Also, subsequent incorporation of a large mass storage system would be eased by the preexistence of a staging computer.

There are two attributes of the staging computer concept which must be weighed against the data characteristics of each potential mass storage application:

- All data sets or files should be relocatable, i.e., must not contain absolute location-dependent data in internal fields.

- If the size of any given file exceeds the total capacity of the disk device, the file must be broken into segments.

On the other hand, a significant advantage of the staging computer concept is achieved when the file transfer operations occur before and after program execution. This enables the staging computer to manage and allocate its own resources independent of the local requirements of an executing program. Conflicting data requests among users can also be coordinated in a multi-tasking, multi-processing environment. Hundreds of files can be transferred to one or more hosts or satellites and made available for processing by programs executing simultaneously. The staging computer hardware resources for performing file transfers can be minimized by using queuing techniques and by time-sharing the hardware data path between a collection of shared peripherals.

Hence, the staging computer is a most versatile approach, gracefully supporting existing applications and new ones in a multi-tasking, multi-processing environment.

5.3.4 Additional Technical Characteristics. Some of the more detailed considerations that would require consideration during a detailed system design would include the following:

- Accuracy and Validity. A primary intertie requirement is the application of effective error detection procedures during each data transfer. Appropriate techniques must be selected to insure that each transfer step is checked for errors. The basic rule is that at least three error checks should be applied to each transmission block, during its transit through the bus. These three checks occur on data leaving a port and data entering a port. There are also checks for transient errors occurring within port processors.

- Timing and Bandwidth. The intertie bus can be as fast as necessary. The objective is to achieve a throughput capability in excess of the aggregate bandwidth of all subscribers. The two variables which determine bus bandwidth are the width of the data base and the speed of the bus drive electronics.

- Bus Acquisition. A high port count in a bus structure tends to increase the probability that more than one port will request bus access at the same time. Throughput demands on the bus require that such contentions are subject to a form of positive control which operates in parallel with bus activity. For this reason, each port must be connected to a bus priority resolver. This module would process bus requests from ports and deliver bus grants to ports contending for bus services. Such operations should occur in parallel with actual bus activity and not affect the overall throughput potential of the bus.

- Global Multiplexing. At any given time several subsystems may be interacting with a host system. The bus is capable of coordinating this activity and the staging computer with its storage capability provides a system-scale buffer.

- Data Routing. Either to or from the central staging center, the system will be capable of performing header analysis in support of routing determination to specific ports.

- Data Translation. Hosts, terminals, and external links need only receive or transmit data in their native code set since the microprocessors will be capable of translating all data that passes through to a normalized code set of the staging computer or back-end processor.

- Data Compaction/Explosion/Fixed Field Exclusion. Microprocessors have the potential of reducing central processing overhead as well as reducing network bandwidth requirements. In this sense the intertie increases effective available bandwidth by insuring certain efficiency procedures.

- Global Access. Valid files may be communicated between any two ports. Resources may be accessed from any valid processor.

- Security. Since all intersubscriber data is first seen by the port processor, it can prevent unauthorized routing of data.

- Input/Outputs. Data transiting the intertie media have three basic source and/or destinations: the central host systems, dedicated I/O systems, and the back-end processor/staging area. The staging computer/back-end processor is the controlling node. That is, most data transits go to or from the staging area as an intermediate source/destination.

- Data Characteristics. The bus should be capable of handling a variety of data types with a fixed protocol. Established data characteristics on the bus include:

-- Fixed transmission block size
-- Byte parallel (8 or 16 bits)
-- Asynchronous

5.3.5 Security Considerations. A primary consideration in the design of both the network and intertie concepts is the element of security. Any techniques to interface computers electronically must provide the necessary capacity to preserve differences in security levels between systems. Further, any such computer intertie techniques must comply with appropriate local and DOD security regulations pertaining to safeguarding classified material in ADP operations. The conventional security system is maintained within the proposed architectural concept of interconnection via a system bus and microprocessor controllers.

5.3.5.1 Logical Security. All data entering the bus is subject to analysis by an isolated processing entity, the port processor, prior to being routed to its destination. Such checking means that the level of logical security protection is equal to that of its subscribers; that is, the port processor will support and not violate any established system privacy rule of which it is cognizant. Since control programs for the bus are ROM contained, established security rules cannot be modified by external sources remotely, i.e., modification of ROM-contained programs requires physical penetration of the system.

The above security rationale means:

- The bus concept is uniquely suited to support system security procedures.
- The port processors make decisions before data is actually allowed to be transmitted to any location.
- The microprocessors are capable of performing many such evaluations in parallel without impacting any subscriber resource.
- The ROM program can allow the dynamic modification of security rules, i.e., altering rules to fit certain changeable sensitive situations.
- The design can provide for encryption/decryption within ROM-contained port programs if security procedures permit.

Encryption/decryption within the microprocessor ports would make the port programs highly classified. Other security procedures within a port, such as domain processing, can be implemented without classifying these programs, i.e., knowing such rules does not readily allow the security of the system to be compromised.

With respect to an integrated data base, on a distributed network, security classification identifiers can exist at the record level or even at the element level of the data structure and all transfer of data between network

"nodes" will be checked to ensure the receiving terminal, processor, or communications line is authorized access to the information and is cleared to receive it. The actual checking of authorizations can be accomplished via system software and by firmware built into the port processor.

Given appropriate implementation of these techniques and possible changes in security regulations, the possibility that implementation of the intertie concept will encounter major problems with respect to full compliance with security regulations for ADP systems is greatly reduced.

5.3.5.2 Physical Security. Two factors impact the physical security of intertie:

- Physical security of logical security enforcement procedures (ROM memory)
- Physical security of the harness which interconnects the subscribers

The first factor is self-explanatory; the second item, however, is more difficult to deal with. The physical harness design must be of top priority in any subsequent system design effort. The existing plant facilities of the DMATC are well suited to insure adequate physical security of critical components of the system.

5.3.6 Software Considerations. Implementation of a full network concept will involve extensive software development. The first interface of any two pieces of dissimilar hardware almost always involves some new software development. Software considerations for the intertie concept are involved at two distinct levels--the microprocessor port processor and the subsystem interfaces to the network.

5.3.6.1 Port Software. The microprocessor-based bus requires the existence of a unique software generation system. Such a system must be capable of assembling firmware for the microprocessors which interface with the staging processor. Such a system also requires provisions for loading assembled programs into read only memory (ROM). Operational software is normally contained in ROM and so requires specialized startup procedures.

It should be pointed out that a major advantage of the microprocessor approach is the relatively small software effort that will be required for its implementation. This is true, since individual port software has independent processing resources. Hence, no complex executive type software is required, i.e., each port has its own memory so no memory manager is required. Each element has its own processor so no queue management is required. Likewise, task scheduling is not required. Such software is normally complex and requires extensive debugging due to its re-entrant nature. The elimination of such software represents a major cost advantage of this approach.

5.3.6.2 Other Software Considerations. The software impact of augmenting the present central processors would be mainly that of continuing to improve, expand, and streamline present programming efforts. If options were

selected which provided the direct interfacing of an associative array processor or a mass storage device to a central processor, a unique interface requirement would be faced. To date neither of these types of equipment have been interfaced with UNIVAC mainframe processors.

In addition to the software interface which is required when two types of hardware are married for the first time, implementation of a Mass Storage System or an associative array processor would require specialized software consideration within the framwork of the intertie concept.

5.3.6.2.1 Mass Storage Systems. Adding a MSS to the system would generate software implications of the following types:

- Programming considerations--Programming that must be considered by the application programmer when creating and requesting data set and system programming necessary for installing MSS software
  - -- The primary user programming consideration for executing jobs in the MSS environment relate to job control language and the direct access support needed by the application. All data sets controlled by MSS must be treated as disk data sets stored by the application programs. Access to application data sets is via existing device support software for disk storage. Applications that use device-independent coding techniques can move their data sets from tape storage to disk by changing the JCL. With the use of translators in a staging computer few software alterations should be necessary.
  - -- System programming considerations include system generation on the MSS hardware, MSS system modification, and I/O device and SYSGEN options on the staging computer.

- MSS monitoring--The MSS would maintain an historical record of system activity. The history would be a chronology of significant system events. The objectives of event monitoring and recording would be to provide:
  - -- Accounting information
  - -- Performance data
  - -- Error recording
  - -- Security data
  - -- System activity

- System maintenance and diagnostics--Support which would consist of a diagnostic monitor and a collection of test programs. These would be used to detect hardware malfunctions and aid the field engineer in the checkout and maintenance of the MSS equipment. The test programs would be started by the monitor and would operate in either online or stand-alone mode.

5.3.6.2.2 Associative Array Processors. Although any new system may present new or different languages to the applications or system programs, this may be a major concern with the AAP. It should also be understood that

associative processors (such as STARAN) are special purpose. They are useful for a limited set of applications which may be characterized as having:

- A large number of independent data sets
- The requirement for quick throughput response
- The potential to exploit associative addressing selection techniques

Since associative machines are special purpose, software implications vary with the application. There are, however, some visible software trade-offs. Higher order language extensions have been considered for use on associative machines. The major extensions of the languages would likely include:

- Data declarations--A means of differentiating processing element and control unit variables is necessary.
- Arithmetic and logical expressions--The use of arithmetic and logical expression is usually directly extended.
- Control statements--The control of the program in an associative machine involves not only the type of control statements seen in conventional machines, but also control statements that use amount of activity as a means of execution control.

5.3.6.2.3 Microprocessors. Present trends indicate that it is becoming increasingly cost-effective to accomplish some kinds of widely used and repetitive functions via specialized logic or microprocessor firmware rather than with generalized computer software. The one-time purchase cost of such hardware may be more economical than the ongoing costs of software development. Thus, the intertie with its microprocessor ports could potentially have an additional advantage of reducing software development efforts.

5.3.6.2.4 Applications Programs. Centralized systems encourage standardization, reduction of redundant operations, and easier quality control. In a distributed system, on the other hand, there is greater user participation (and hence, satisfaction). In general, the nature of applications programming will be similar regardless of the implementation route chosen. During the system development, and as current production trends may reach hardware limitations, it would be wise to work out interim stopgap solutions, such as streamlining of software and program libraries.

5.3.7 System Comparison. At this point it may be of value to compare existing methods and procedures with the type of system envisioned in the intertie concept. After briefly comparing the "now" and "then," some examples of data flow under the new configuration will be used to illustrate typical information paths.

5.3.7.1 Existing Methods and Procedures. The DMATC computer complex today consists primarily of interdependent but isolated processors which tend

to be single processing entities surrounding two large central host computers. While one processor such as the Site I computer may require data generated by CELESTE, that interchange is now accommodated manually by human intervention. This human intervention is time consuming and introduces an error-prone activity in the tape handling component. Also, since in a given time only a limited number of manual operations are practical, central computer data interchanges tend to be large; whole files are interchanged rather than run the risk of not satisfying demands. This bulk interchange also tends to increase the size and speed required by central computers. Human intervention in inter-data transfers, in other words, slows things down and imposes artificial cost burdens.

5.3.7.2 Proposed Methods and Procedures. The proposed DMATC intertie network envisioned could be characterized as a microprocessor based bus. The bus will act as a master communication channel for all intersubscriber traffic. Primary responsibilities of the microprocessors include:

- Routing determination for all intersubscriber data
- Message or transmission block queuing for all subscribers
- Protocol enforcement for all subscribers
- Interface normalization as required
- Error checking for all transfers
- Message or block multiplexing services
- Data compression and explosion
- Data translation services as required
- Network control services
- Enforcement of logical security rules on network-wide basis
- Error sealing services, i.e., transient errors can be "sealed off" before generating system-wide impact

The above capabilities combine to support the capabilities listed in the previous section.

A major benefit of the bus and microprocessor approach is the capability for implementation of high-resolution system-to-system data selectivity. The current off-line procedures for system-to-system communication require the transfer of large data files via tape, primarily due to historical developments. The sending system generally sends a full file from which the receiving system may process all or only a subset of the records or fields. Thus, all data that has historically been generated is transferred. The presence of a

universal data communications link between these systems could allow the recipient to access only those records it needs through the staging computer. This automatic selectivity of data can reduce network traffic and improve responsiveness by reducing the amount of transmitted data and by eliminating the human element in the tape transfer process.

5.3.7.3 Scenarios for Job Flow Through New Configuration. The intertie concept provides such a flexible means of tying data processing components together that one can postulate an almost endless variety of information paths. A few of these are suggested below.

Scenarios for Application Programming. Figure 23 shows two of many possible program development job flow sequences. The circled numbers in the figure indicate (1) entry or editing of program instructions, (2) assembly or compilation, (3) testing, and (4) storage in a program library.

Figure 23A depicts the situation of an intelligent satellite processor. Object program creation, testing, and storage are done locally, with optional object program storage also at the central facility.

Figure 23B illustrates a case where the equipment at the satellite location is less elaborate, perhaps a CRT terminal. Only entry and testing are done at the satellite.

Data Movement to a Satellite. Figure 24 illustrates a typical data movement sequence corresponding by number to the following. A query (1) is entered at a distributed site. The query (2) is submitted to the data base in mass storage (4) via the staging computer (3). The staging computer (5) moves the requested data back to the query originator (6). If the amount of data to be moved is large, it may temporarily reside on a staging disk (not shown) on the way to the originator.

Data Base Update. Data base update flow is illustrated in Figure 25 showing the following numbered steps. Update information (1) is entered through the central computer, which communicates with the staging computer (2). The records to be updated are moved from mass storage (3) to a staging disk (4). The update occurs and the changed records are returned to the data base storage (5). A transaction record is returned to the point of origination (6). Not shown is an administration terminal linked directly to the staging computer through which data base management processes are controlled. This terminal might also be the recipient of the transaction record.

Balancing Loading Assignments Between Hosts. Figure 26 shows how incoming jobs can be assigned to the least busy host by treating request queues and job files as additional files at the staging computer. A job (1) comes in through Host A. The appropriate program file is accessed from mass storage (2), job queues are checked (3), and the least busy appropriate host is selected for the actual processing (4). The results are returned, if desired to the originating computer (6); and request queues are updated (5).

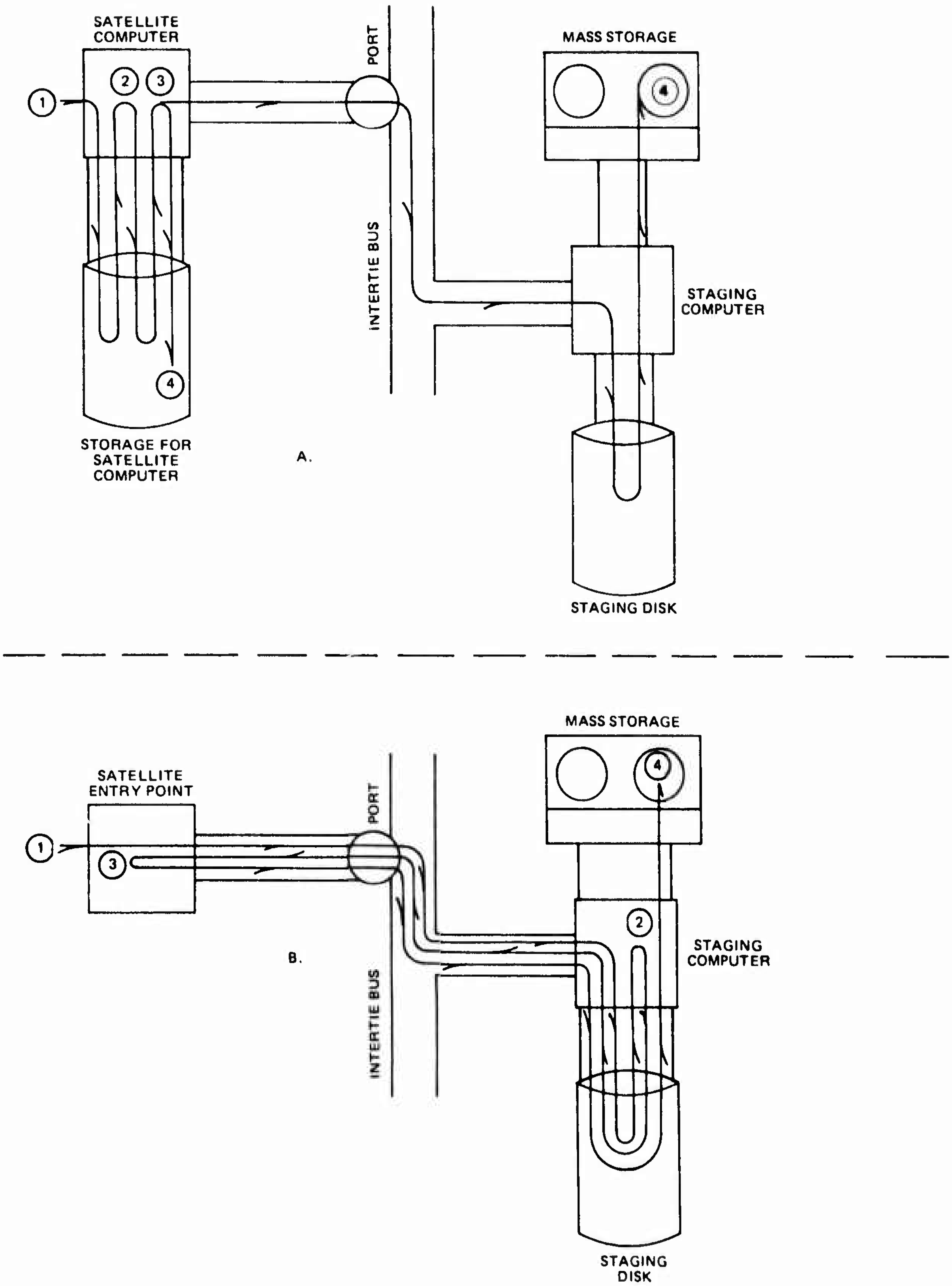


Figure 23. Possible Job Flow Scenarios for Application Programming

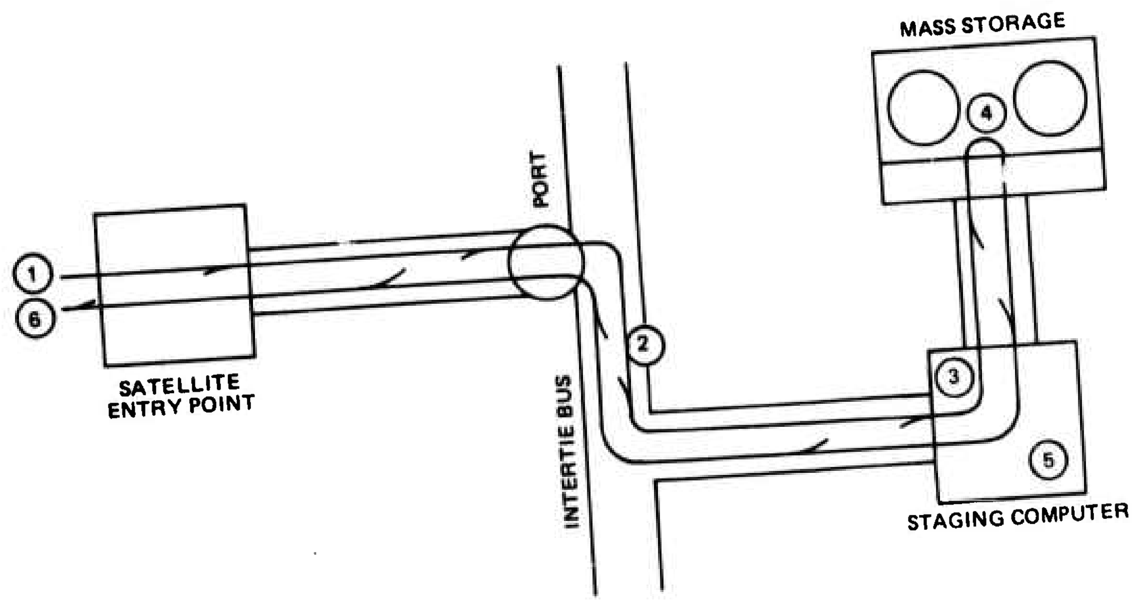


Figure 24. Job Flow–Data Movement to a Satellite

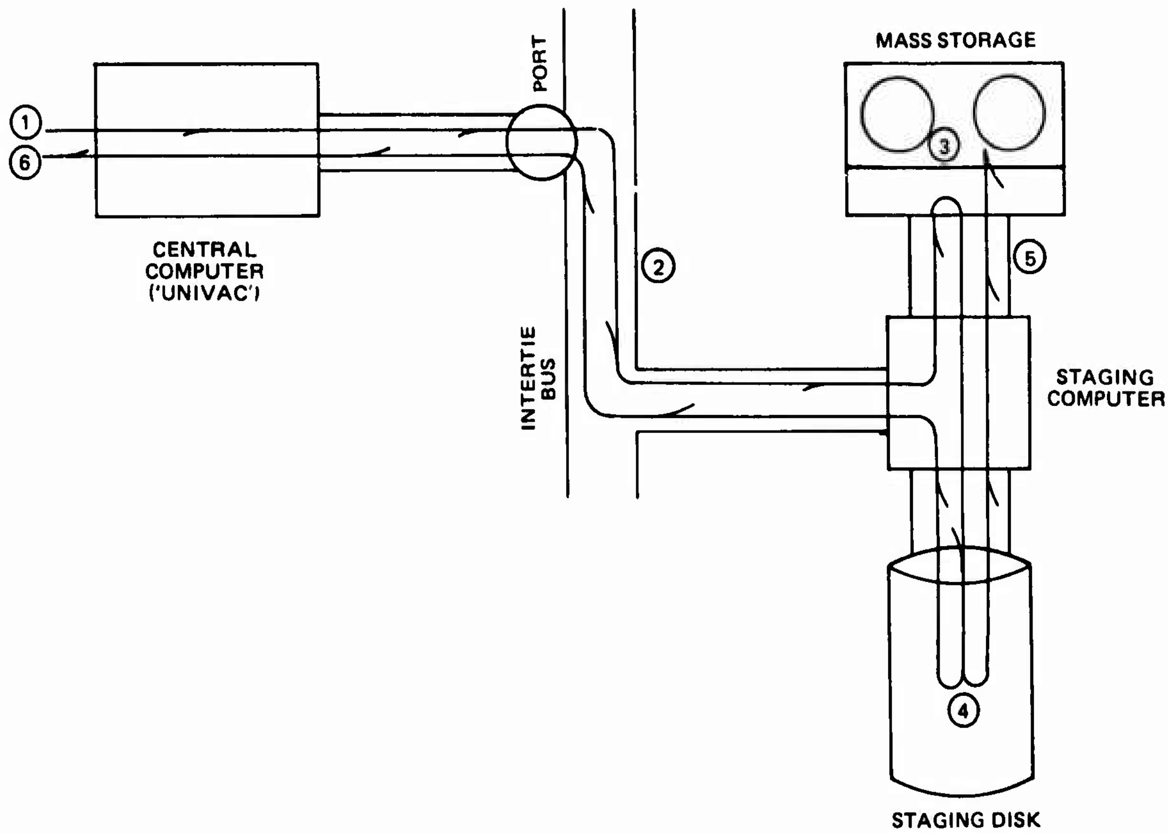


Figure 25. Job Flow–Data Base Update

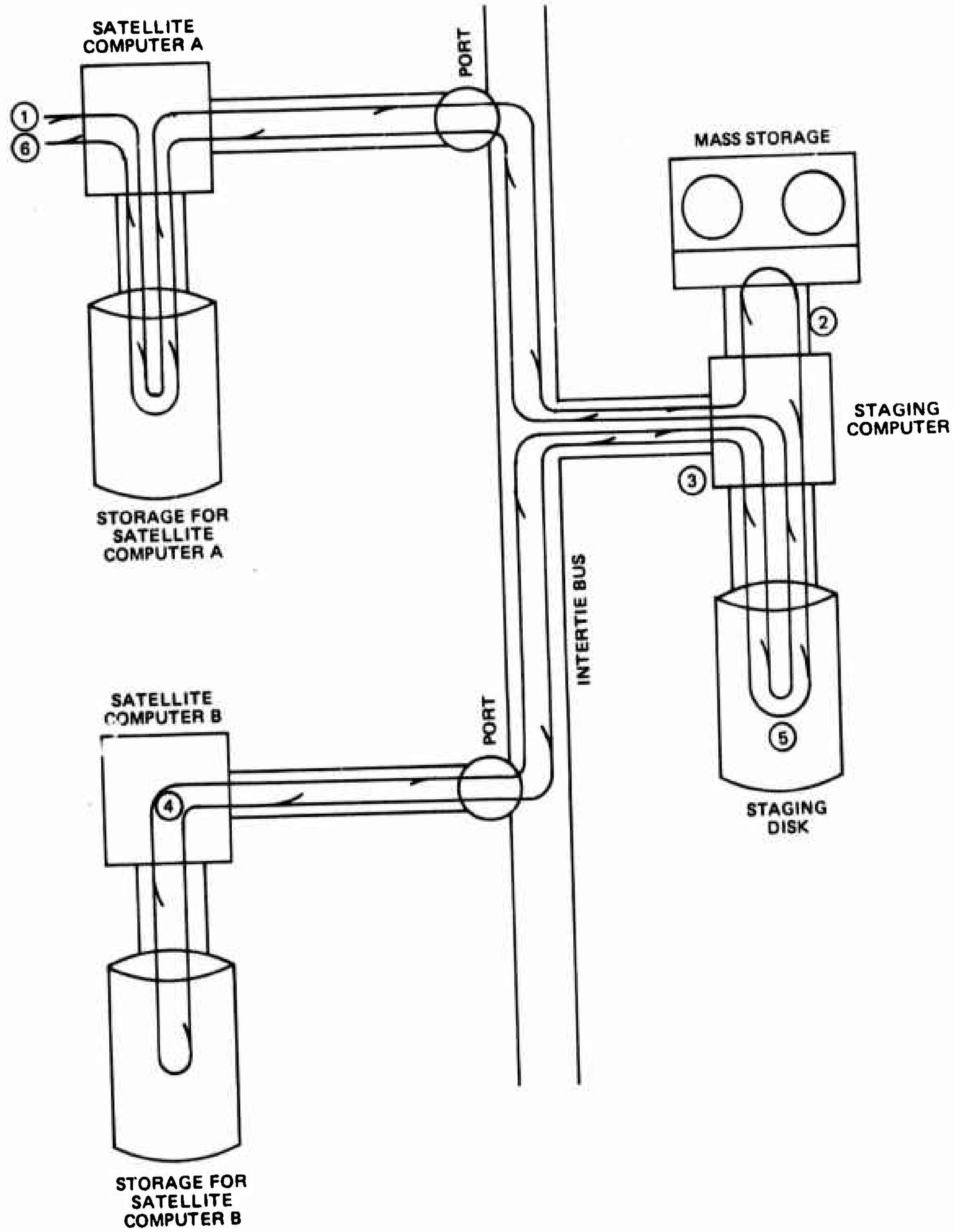


Figure 26. Job Flow—Balancing Loading Assignments

5.3.8 Configuration Options. Numerous variations could be proposed on the theme which has been described in the preceding sections. A sampling of such options is listed below. A careful look should be taken at the trade-offs involved in each before arriving at a judgment of what constitutes the optimal configuration.

- The staging computer or back-end processor may or may not be the same hardware as the mass storage system controller.
- The mass storage system itself is entirely optional. Its level of performance will largely depend on the nature of the data stored and DMATC response requirements.
- If required, a mainframe processor can be isolated from the rest of the system for secure processing.
- An associative processor (of the STARAN or other type) could be attached to the bus or to a host computer mainframe.
- Most other bus connections are optional. The existing tape interface scheme could be selectively retained based on subsystem functions and activity.
- Complete backup capability would be provided if both host computer mainframes had a full range of identical peripherals.
- There is great latitude in the arrangement of disks, tapes, and the attachment of satellites.
- The staging computer could possess a degree of complexity anywhere in the range from a simple buffer processor to a back-end processor for an elaborate data base management system.

5.4 Implementation Sequence. The implementation of an advanced architecture must be achieved incrementally. Figure 27 shows a possible implementation sequence for hardware components of the system. For the necessary events to be realized, certain plans and strategies would have to be initiated early in the implementation process due to the long lead time factors for subsystem development and equipment procurement.

A detailed trade-off analysis will be needed to select a specific system configuration. This will entail a detailed analysis of the trade-offs involved in each option before arriving at a judgment as to the optimal configuration. For example, the manner in which a Data Base Management System is to be addressed has a significant design impact. Recognition of various options and alternatives will place an added emphasis on strong, centralized management and planning to insure an integrated and cohesive capability to meet production requirements in the 1980's.

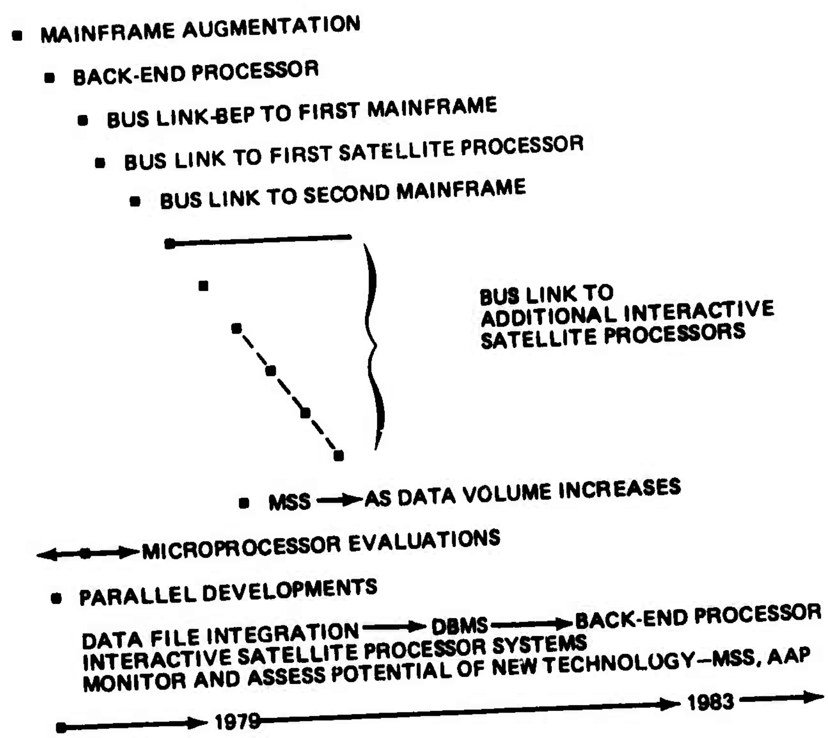


Figure 27. Implementation Sequence

Ample planning time for each component also allows time to study the actual impacts of each on the whole processing environment. The schedule is not as important as the order of implementation.

DMATC stated requirements show that the existing mainframe processors require an early augmentation. Thus, the first order of business is to consider increased central processor capabilities. As the work load increases on the augmented central processors, a corresponding need for increased I/O capacity will have to be addressed, as data interfacing volume increases.

Concurrent with the augmentation of the central processors, parallel efforts should be initiated to further define other system objectives and functions. These efforts will include:

- Definition of an integrated data base management function
- Development or augmentation of satellite processing systems addressing high activity functional areas such as:
  - -- Cartographic Interactive Data Edit
  - -- Photogrammetic Postprocessing
  - -- CELESTE Processing
- Further definition of network intertie functions and components

Probably the one component in the entire configuration whose design and implementation will require the longest lead time is the network intertie. Full system definition, development, and testing could involve as much as four years lead time. It should have a priority immediately after central processor augmentation. Similar efforts currently under development should demonstrate the technical viability of the intertie concept and significantly reduce technical risk.

Definition of the role of the data base management function will involve a systematic analysis of all data files and holdings to identify the optimum means of supporting an integrated data base function. Regardless of the intertie concept, this effort should be pursued on a priority basis to provide more effective management and utilization of data holdings. Assuming that this analysis will support adaptation of a Data Base Management System, the further definition of how the data management function would be accommodated in a network concept is a basic consideration. From a conceptual point of view, however, pursuit of the back-end data base management (BEP) approach is considered the most desirable approach. In addition to helping to define a data management strategy for the future, a comprehensive program addressing restructuring of data holdings will provide a stronger basis on which to define selection and implementation strategy for integration of a mass storage system in later phases of the system development.

The development of dedicated satellite processing systems, particularly as related to the creation of error-free digital records from cartographic sources, also represents an area which should be pursued on its own merits. Improved methods of creating and editing digital files and minimizing the throughput limitations associated with batch processing will be required independent of

a decision to enter into a network configuration. The design of the intertie concept allows interfacing with these dedicated subsystems, but is not constrained by the subsystems themselves.

A similar development effort addresses photogrammetric edit and postprocessing functions. Using a conceptual approach similar to the Integrated Photogrammetric Processing Network (IPIN), a system reflecting DMATC-planned photogrammetric capabilities and production requirements should be addressed to improve system throughput.

Augmentation of existing CELESTE capabilities would provide a third logical area in which distribution of the processing work load would facilitate information flow and improve production throughput.

The evolution of DMATC processing environment into a highly integrated capability will require parallel action in the above areas. The ar as cited are considered basic to an orderly growth of DMATC capabilities regardless of the advanced system concept selected. In addition to these actions, advances in technology related to mass storage systems and associative array processors should be continuously monitored. In both of these areas emerging technology holds good prospects for increasing the viable options which may provide cost-effective solutions to projected DMATC needs. The important factor is that each of these potential system components should be viewed from the "top down" and in terms of its specific contribution to the Center's long-term objectives.

Beyond these, it would not be wise to lock into a particular physical installation scheme based on today's technology for a system whose operational target is in the 1980's.

The realization of objectives by the 1983 time frame therefore requires the following projects:

- Augmented central computer systems
- Development of interactive satellite processing subsystems
- Intertie design and implementation
- Development of a data base management concept and acquisition of a suitable Data Base Management System (potentially a back-end processor system)
- Possible acquisition of a mass storage system

If the above assumptions represent a reasonable scenario, it is believed that the funding should be allocated in the same sequence. Some phasing at any rate is necessary to prevent funding peaks that would require major resolution problems.

Finally, as the system concept presented reaches maturity and significant portions of the processing work load are distributed, the nature and capacity of the mainframe processors will require re-evaluation. While specialized functions will be offloaded to the dedicated subsystems, it is reasonable to expect that a significant general work load will still require support from a mainframe processor. Even if a dramatic offloading were achieved, it would not occur until late in the time frame addressed in this study.

5.5 Applicability of the Intertie Concept to DMAAC. The computer intertie concepts proposed in this section are considered equally applicable to the DMATC and the DMA Aerospace Center (DMAAC). In both cases the evolution and demonstration of back-end data base management could provide a viable option to support an integrated data base management function. In either installation, adaptation of the concept would involve a transition toward online systems and online data bases. The trend toward functionally-oriented data bases, such as the Automated Aeronautical Information Processing System (AAIPS) or a Gravity Data Base at the DMAAC could be readily accommodated within this concept. In both cases, some level of computer interface is inherent as it is in any truly distributive processing concept. Similarly the same constraints, including an accommodation of multi-level security requirements, are the most severe hindrances to evolution of an integrated system concept. DMAAC projected data processing requirements identified in July 1975 reflected a stable requirement for SAA processing capability. If this projection remains valid, security constraints may in fact be less significant for the DMAAC and allow earlier integration of a significant portion of the DMAAC data processing environment.

6. CONCLUSIONS

This section presents a summary of the most significant conclusions reached during the study effort. Although the conclusions are presented as individual items, there is a high degree of interdependence between both the conclusions and individual recommendations which follow.

6.1 Central Processing. Based on the requirements provided by DMATC, a major augmentation of the central processing capabilities will be required in FY 1977.

- Projected growth in processing requirements at the SAA and the collateral levels will require augmentation at both of these security levels.

- As the work load increases on the augmented central processors, a corresponding need for increased input/output capability can be anticipated.

- As the level of processing activity increases, use of magnetic tape as the primary file transfer media will increase tape proliferation. Quality control problems associated with manual handling of large quantities of tape can be expected to increase dramatically.

- Continued operation in a batch mode for processes which require numerous sequential processing steps (such as cartographic data capture and edit) will limit the reduction of production throughput time.

6.2 Dedicated Processing Subsystems. Improved production throughput time can be achieved by a transition toward distributed satellite processing systems. These dedicated systems would perform data edit functions in an interactive mode and support routine data functions.

- Based on an indepth examination of the SACARTS system, a significant improvement in the calendar time required to generate digital cartographic files will require implementation of an interactive cartographic data edit system.

- Development of a photogrammetric postprocessing system, similar in intent to the Integrated Photogrammetric Instrumentation Network (IPIN) would allow performance of edit and formatting functions at a rate consistent with the output of photogrammetric devices. Design of this system should be tailored to DMATC photogrammetric capabilities and production requirements.

- Continued expansion of the capabilities within the CELESTE processing system to support data edit/formatting functions would minimize the impact of the high input volume seasonal work load on the central systems.

- Implementation of dedicated processing systems will involve extended lead times. Consequently, their potential for offloading processing from the central systems would not be realized for several years. The extent to which these dedicated systems will relieve the central processor work load will depend on the functional capabilities provided. Their most dramatic effect, however, is expected to be improved throughput calendar time.

6.3 Data Systems. From a functional view, the work load and data storage associated with the production of 1:50,000 scale topographic maps will remain the dominant factor in the DMATC through the 1983 time frame. (Attention is drawn to the Addendum to this report which cites a recent change in DMATC program direction.)

- The SACARTS system, both in terms of its hardware components and the GIST software system, should receive priority attention to improve its capability to support this primary production area. Continued actions to improve both throughput capacity and throughput time is essential.

- Data files constituting a potential source data base (UNAMACE masters, SACARTS, DTIB) could reach a level of 4.7×10^{12} bits by 1983. This estimate is based on projected magnetic tape holdings and should be subjected to further verification. Within the context of this study these files are considered primarily archival in nature.

- Projections for online data storage indicate a potential requirement for approximately 225 million words, exclusive of catalogue file and scratch space. This estimate reflects the study team's assessment of minimum online file capacity, rather than user stated turn around requirements. A continued effort to define this requirement based on production response requirements is warranted.

6.4 Information Management. Data files currently maintained by DMATC are essentially independent files. In most cases, interactions between files must be manually generated. This mode of information handling precludes the rapid exploitation of interrelated but independent fact files to responsively support management functions.

A need is recognized for an improved file management capability to provide more responsive support to the production planning process. This need will require a systematic approach to allow the integration or enhanced responsiveness of files which identify the availability of source materials suitable for the support of new production requirements.

6.5 Security. Protection of multi-level security information is and will remain a major consideration and constraint during the evolution of any DMATC system architecture.

6.6 Advanced Technology. While virtually all aspects of data processing technology are changing at a rapid pace, the following represents the study teams view of how specific technologies pertinent to the study are expected to affect the evolving DMATC environment.

6.6.1 Mass Storage Systems. Prospects for significant breakthroughs in mass storage technology are not encouraging. While advanced methods will mature, costs will remain high. Costs associated with new technology are expected to remain high as compared to magnetic tape technology. Significant growth potential in magnetic tape technology will tend to maintain the relative cost advantage of tape technology for some time.

6.6.2 Associative Array Processors. Despite demonstration of cartographic applications on the associative array processor, extensive research and development is required to provide a system suitable for use in a production environment. Major development efforts are still required in areas of mass storage and high order language development.

6.6.3 Microprocessor Technology. Microprocessor technology is expected to become increasingly important in the performance of high-volume repetitive processing requirements associated with cartographic production. The potential use of microprocessors to perform a control function is also expected to play a major role in providing increased security control within integrated data management structures.

6.6.4 Back-End Data Base Management. The utilization of minicomputers to support a Data Base Management System is a promising technology providing a broad scope of architectural options. A principal motivation for this type of data base management system is the alignment of processing functions with machine capabilities and hence maximizing the efficiency of applications-oriented processors such as the UNIVAC mainframes. Back-end data base management holds promise for an economical solution to data base management and can be adapted either for completely centralized data base management or a distributed data base concept.

7. RECOMMENDATIONS

The following recommendations are limited to the major actions considered essential to meet the data processing and data handling requirements of the DMATC in the 1977-1983 time frame. No single recommendation is considered adequate to provide a system capability which addresses all facets of the increased processing and data handling problems facing the Center. Consequently, the following parallel actions are recommended to provide the balance efforts considered necessary to meet DMATC goals through 1983 and beyond.

- Continue actions to augment the central processing capability of the Center. The processing requirements as provided by the DMATC support augmentation with a UNIVAC 1100/40 or equivalent processor.

- Initiate actions to provide an interactive cartographic data edit subsystem to improve production throughput time for creation of digital cartographic data files. The system selection should be a design that can be easily adapted to meet interactive edit requirements of the Field Offices.

- Initiate actions to define functional and system requirements for a photogrammetric postprocessor tailored to DMATC production needs.

- Continue actions addressing information management to allow more responsive exploitation of resource files. Achievement of this objective will require the following actions:
  - -- Examine files with a view toward integration or interfacing of related fact files
  - -- Examine the feasibility of structuring fact files under a Data Base Management System
  - -- Continue file evaluation program to refine data volume estimates for online storage and archival storage requirements
  - -- Continue file evaluation program to refine file access and response time requirements to provide planning data for refinement of mass storage system requirements

- Continue efforts to improve production throughput capacity of the SACARTS system. Specific actions should include:
  - -- Refine the role of the scanner/plotter and DIODE system in the production cycle
  - -- Initiate efforts to examine alternative options for scanner/plotter data processing support
  - -- Consider development of an intermediate data structure for GIST levels 1 through 4 to support online interactive data capture/edit functions
  - -- Plan for the integration of automated cartographic functions in the Field Offices within the scope of SACARTS production activities

- Initiate research and development programs supporting an evolution toward an integrated system architecture as proposed in the computer intertie concept

- Integrate within the overall ADP support plan actions addressing other Center processors. Because of their age replacement of both the B-3500 and the IBM 7094 must be anticipated. Replacement of the Geodetic Survey Squadron hardware should be addressed early because of potential costs associated with old equipment. A requirements definition study should be initiated with replacement by a minicomputer as a prime consideration.

APPENDIX A

SACARTS

A.1 Introduction. Because of the heavy production work load programmed for support by the Semi-Automatic Cartographic System (SACARTS), the need for a detailed examination of this process was recognized early in the study. As a major user of UNIVAC 1108 resources, increased production of topographic maps to be supported by SACARTS to a level of 500 map sheets per year was recognized as one of the principal DMATC growth factors.

As the effort progressed, recognition of the relationship of SACARTS to other programs (such as the raster scanner plotter, the DIODES system, the potential entry of an ADP capability in the Field Offices) and the expected role of automated cartography in long-range PDOP considerations, emphasized the significance of SACARTS and the importance of resolving related problems.

A.2 Background. The development and application of automation support to the cartographic production requirements of the Topographic Center (TC) has received increasing emphasis over the past years. The application of this technology has followed the general pattern set by industry at large. That is, it has progressed from accounting-type functions, through the scientific and technical area, into production management and information storage and retrieval applications. Paralleling industrial developments, the application of automation technology to cartography began on the periphery of the process and is working its way toward direct and more efficient support of the overall production activity.

The impact of technology on the forms of information input to the cartographic process and on cartographic products themselves has been dramatic. New demands for speed, precision, and product form flexibility are being levied on the cartographer--demands that cannot be met without the deliberate development of new techniques for storing and handling cartographic information. Cartography is a unique and difficult application area for automation; it stubbornly refuses to yield totally to quantitative methods. An extensive, dynamic, scientific method is involved in the practice of modern cartography. However, a generous portion of the art of automatic cartography remains unrevealed; this promises to keep the cartographer very much involved for years to come. This is significant to the designer and builder of cartographic support systems since, to be successful, such systems must provide for purposeful human interaction at strategic points throughout the process.

To date, three major automation development efforts have focused on several aspects of the production process: the Digital Topographic Data Collection System (DTDCS); the Universal Automatic Map Compilation Equipment (UNAMACE) System; and the Semi-Automatic Cartographic System (SACARTS).

During the development of the above three systems a series of data files was developed for the direct support of cartographic maintenance, production research, cartographic compilation, and topographic production functions. Experience gained through the development of these systems has resulted not only

in a mature understanding of the cartographic products, processes, and information requirements, but an intimate knowledge of the production methods used at the Topographic Center.

A.3 SACARTS Development. The development of SACARTS at the Topographic Center is one of the major efforts concerned with computer-supported cartographic production. SACARTS was developed in a quasi-experimental environment, one that addressed both the hardware and software aspects of an automated cartography system. Work station development resulted in the combination of state-of-the-art and industrial "shelf" items yielding several types of digitizing subsystems (analog-to-digital) for the capturing of raw digital data and for providing a means by which previously digitized data may be modified and edited.

Extensive investigation by DMATC established that the software required to support "auto-carto" requirements did not exist. Many of the problems had not been addressed in depth while others did not preserve a system concept and/or maintain the required accuracy. It was decided to design a total software system.

The Graphic Improvement Software Transformation system (GIST) is the product of that design. It is a batch-oriented modular system that operates on the UNIVAC 1108. GIST consists of seven modules which can produce a series of color-separated plots employing a limited repertoire of symbolization operations. These plots are suitable for producing a finished reproducible product.

A.4 Identification of Problem Areas. The review of procedures and techniques related to the use of SACARTS was conducted by interview with working level personnel using the system, by familiarization with actual operations, and by examination of software documentation for the GIST system. During the course of the review, three problems were identified as significant which limit the throughput capacity of the system or cause undesirable burden on processing facilities. Two of these problems are associated with the GIST software system--namely, the use of a batch mode operation during the data capture phase, and the efficiency of data base access techniques. The third problem area is associated with the SACARTS subsystems used during the digitizing and editing process. Although these problems are not totally independent, they will be treated as three separate topics.

A.4.1 GIST Batch Mode Operation. The GIST system is a collection of program modules used to create and maintain a digital data structure in the automatic cartographic effort at the Topographic Center. Each module executes on the UNIVAC 1108 in an independent fashion and in a batch mode environment. While each module appears to perform its function as originally intended, use of the batch mode of operation constitutes a major drawback to efficient high volume production.

Figure A-1 is a generalized production flow diagram showing the SACARTS processing steps and illustrating the GIST modules. A significant point that becomes apparent in the diagram is the requirement for human intervention at each step in the processing sequence. The efficiency of the SACARTS depends

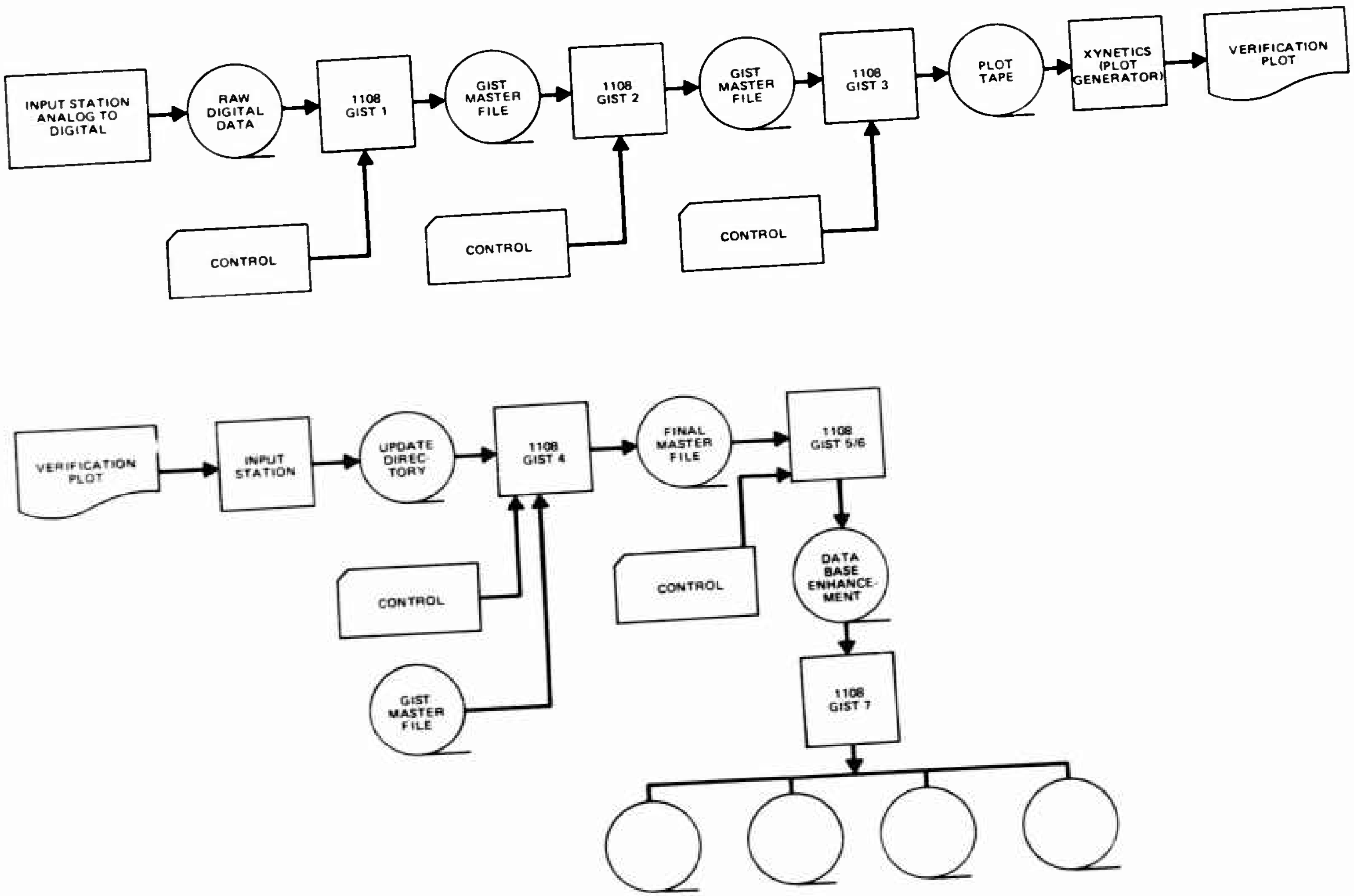


Figure A-1. SACARTS Functional Flow

in large measure on the degrading effect of the man-machine/human-intervention requirements of the batch modules of the GIST software package. In essence the original objective of minimizing human intervention in the GIST process has been thwarted by the manual intervention imposed by the existing batch mode of processing.

The modular segmentation of the GIST functions operating in a batch environment imposes an unreasonable delay, measurable in days, in creating a digital file in which all features are correctly positioned and classified. This is due to need for sequential processing of the GIST modules 1, 2, and 3, and then the iterative cycle of executing GIST modules 4 and 3 to accomplish the edit and verification functions of the data capture phase. The need for GIST module 1 to successfully complete before GIST module 2 can be executed may in itself cause one or more days delay in a batch environment since the execution could be aborted for a variety of reasons such as an error in the GIST 1 input parameter card deck, the correct work tape not found at the UNIVAC 1108, parity on work tape, etc. The possible causes for delay are many, expensive, and unacceptable. Each of the GIST modules requires a human to handle an input tape (at least one), an output tape (at least one), and to prepare and handle the input parameter card deck used to control the specific functions to be performed by the called GIST module. GIST 4, in addition to these restrictions, requires the cartographer to follow a stringent set of procedures while performing edits to a master file. Violation of these procedures could cause an edit update job to a GIST master file to be aborted and/or possibly contaminate the data file.

The segmentation of GIST modules 1 through 4, combined with the UNIVAC 1108 batch environment, may have been an asset in the GIST development phase, but it proves to be a detriment in a production environment. GIST was designed as a passive data base manipulator. There is no direct interplay with the cartographer and the data file and, an excessively long calendar time is required to acquire an error-free digital master file.

In summary, the GIST software performs two general functions: GIST levels 1-4 support the accumulation of an error-free digital file with all features classified correctly and positioned properly. GIST levels 5-7 support the symbolization and color separation of the error-free data in a digital file. The latter function lends itself quite readily to a batch environment and therefore no real loss is suffered through its residence and execution on the UNIVAC 1108; the former, as previously discussed, does in fact suffer greatly in the form of an extended time requirement in performing its tasks.

A.4.2 Data Base Access. A data structure, as detailed in Graphic Improvement Software Transformation by H.R. Cook, was created to support GIST. The data structure incorporated in this design provides access to data to perform local edit in a batch mode. In this mode the speed with which a GIST master file can be created and maintained may not be critical. Data access time becomes critical when use of an interactive data capture and edit system is considered. Rapid recovery and display of data in the interactive mode is a critical human engineering factor. For use in an interactive system it is essential that the data base structure be such that any feature in the digital file can be located quickly and efficiently.

The GIST data structure uses a Cartesian coordinate system which assists in subdividing the surface of the source document (chart, photo, plot, etc.) into a series of well defined areas of half-inch squares called sections. Each section contains a number of "pages" which in turn is comprised of a group (11 maximum) of data nodes. It is these data nodes that contain feature classification information and coordinate (X, Y) information in the form of chain-encoded vectors that describe the convolutions of the feature. Since a node is finite in length, it may be necessary to chain or link a family of n nodes to fully describe a feature. Classification information stored in each node is redundant in all but the first node in this case.

When editing, the cartographer identifies a position (X, Y) on the source document via the stylus location. The UNIVAC 1108 batch GIST module 4 will use the X, Y position to select a data base "section/page/node." Each of the chain-encoded vectors in each node is processed in a filtered search in an attempt to match a point on the feature with the point specified by the stylus X, Y coordinates. If a match is not made, the next node is processed until a match is found or all nodes in the page have been examined. Similarly, subsequent pages associated with the specified section are sequentially examined in an attempt to locate the desired point.

A major problem of the current data base structure is that a family of n nodes, that is used to define a unique feature, cannot be addressed as one complete feature and must be referenced as a group of sequential nodes. There is no common identifier by which a group of nodes constituting a unique feature such as a road, a river, or a railroad may be addressed. This inability to address a "complete feature" on a sector basis hampers several economic software methods that could be used to decrease the data point search/locate time in an edit sequence.

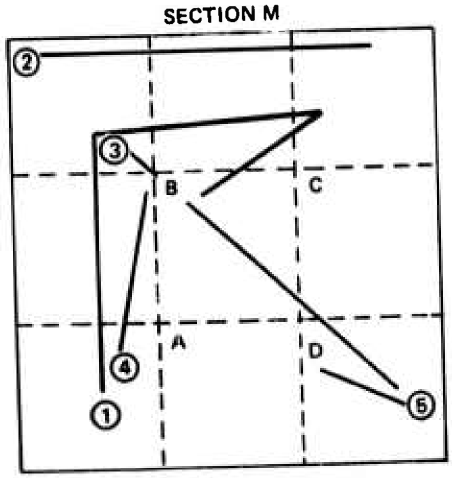
The method which is now used to partition data in a GIST master file may not provide sufficient selectivity to perform an efficient retrieval operation. To illustrate this point we will outline an alternative approach and then compare the data retrieval process using the GIST structure to the alternative.

Assume an alternative approach in which a feature consists of a chain of n nodes and that the chain is identified by a unique name. Assume further that each sector is overlaid with a grid and that each grid intersection is a pointer to an inverse list containing the unique name assigned to each feature that passed through the domain of that grid intersection. Retrieval for edit or examination will be greatly enhanced since the X, Y position of the work station stylus can be used to select the appropriate grid intersect which in turn yields the name of the feature(s) within the domain of that intersection. The random search of data nodes has been greatly reduced if not eliminated.

The following example (Figure A-2) illustrates how the search operation under GIST would differ from the operation using a grid overlay.

The significance of improving the time required for feature retrieval is not as obvious as the consequences of an all batch operation. In the batch mode, increased search time is less apparent, and the processing time may not

SECTION M



GRID INTERSECT TABLE A-1

A	B	C	D
1	1	1	3
4	2	2	5
	3		
	4		

GIST PAGE CONTAINING SECTION M

CONTROL INFORMATION		
NODE	1	F(1)
	2	F(1)
	3	F(1)
	4	F(2)
	5	F(3)
	6	F(3)
	7	F(4)
	8	F(5)
	9	
	10	
NODE	11	

Assume GIST section M with features 1 through 5 in residence. Feature 1 requires 2½ GIST nodes for its description; feature 2, 1 node; feature 3, 1½ nodes; features 4 and 5, 1 node each. Overlay section M with a grid containing 4 intersections A, B, C, and D. Each intersection is a pointer to a list of feature ID numbers showing the features that pass through the domain of the intersection. Table A-1 corresponds with section M.

The cartographer may edit by placing the stylus on any point on the feature of interest, producing an X, Y value for point identification. Feature 5 is selected for modification. The GIST system requires all data in nodes 1 through 8 inclusive be sequentially searched until a position match is made. Using the grid intersect table, the stylus position identifies intersect D as the nearest intersection. Feature identification (3 and 5) are extracted and only 3 nodes instead of 8 will be searched before a match is made.

Figure A-2. Example of an Alternative Retrieval Method

in fact be a major concern. Efficiency of the data retrieval operation becomes significant, however, when use of an interactive edit system is considered. In this situation, the time required to retrieve a feature from the master file will translate to user delay and could constitute a major user frustration. Because of this limitation, and a recognized need to minimize retrieval time in an interactive situation, redesign of the data structure or development of an intermediate data structure specifically designed for interactive file maintenance should be considered.

A.4.3 Data Capture Systems. In this section some problems and possible solutions related to existing data capture/edit devices are addressed.

A.4.3.1 Digital Planimetric Compiler (DPC). This is a special purpose system used to convert source material from an analog to a digital form. The equipment used to achieve this conversion is one mainframe CPU (PDP 11/20), one magnetic tape used as the output medium, and five Bendix Datagrid systems for data capture. The operational procedures required for datagrid setup, data capture, edit, and classification are well known and are not the subject here; instead, some of the DPC system problem areas and their resulting UNIVAC 1108 considerations are the points to be addressed.

- Data Sort Problem. Each work station has a unique identification number associated with it; as the data is created, it is tagged with this identification number and stored in an assigned buffer reserved for that station. The data is written to magnetic tape on cartographer command or when the PDP 11/20 buffer becomes full. The output buffers from each table are written on the output magnetic tape in a mixed random sequence.

- Output Data Format. The output format of the DPC is incremental data and not the absolute (X, Y), CALMA/NOVA format required by the Graphic Improvement Software Transformation (GIST). Therefore, all DPC output tapes must be sorted and reformatted on the UNIVAC 1108 prior to GIST operations. The 1108 sort/reformat sequence may require 11 separate batch 1108 computer runs, the end result of which is five magnetic tapes sorted by work station identification number and in the CALMA/NOVA format required by the GIST system.

- Job Inflexibility. Two different jobs (A and B) may not be sequentially mounted on one work station with both jobs creating data since the table identification alone is inadequate to distinguish job A from job B on the output tape. If the DPC output tape were changed to permit one station to work on two different jobs (A and B), the output from the 1108 sort/reformat operation would be 10 magnetic tapes to handle, bookkeep, and process--when in fact six tapes (two for one station) are all that are really required.

The problems inherent in the operation of the DPC are:

- Table identification and not job identification is used to identify the mixed random data records written on the DPC output magnetic tape.

- Only one composite output tape is generated containing randomly mixed data records from each work station in operation.

- The magnetic tape output format is incremental in nature and is not directly acceptable as input to the UNIVAC 1108 GIST software.

- UNIVAC 1108 time is required to perform the preprocessing sequence of the DPC table sort/data reformat prior to submission to the GIST system.

- The DPC sort/reformat sequence requires a large number of intermediate processes.

- Since the output magnetic tapes generated on the DPC cannot stack multiple data files, each day's work on output tape must be processed on the UNIVAC 1108 yielding relatively small amounts of data per work station and increasing the number of work tapes required.

- Near the end of shift a cartographer may elect not to start a second job on a work station rather than cause an interruption to all other work stations to permit the DPC output tape to be changed.

- Handling a needlessly large number of magnetic tapes increases the possibility of human error, tape error (such as parity), and multiplies the probability of data degradation and loss.

All of the above problems are encountered in the DPC or in the required preprocessing steps prior to entry to GIST module 1. The cycle occurs both when initially creating a file or when updating a master file. The next two points address the GIST operations for each of the data capture systems and not just the DPC.

- GIST modules 1, 2, and 3 must be exercised to create a master file and a proof plot to verify the accuracy of the captured data. UNIVAC 1108 turn around time (measured in days) and the UNIVAC 1108 costs are factors for consideration.

- GIST modules 4 and 3 must be executed repeatedly (on the UNIVAC 1108) to complete the edit and data capture verification cycle. Calendar time and UNIVAC 1108 costs are again the factors that must be evaluated.

When examining the problems listed above, it becomes obvious that the major fault in the operation of the DPC subsystem is the indiscriminate manner in which the output from all work stations is written on the one magnetic tape used as the output device. It becomes equally obvious that this could be the point of attack to make a significant improvement in the operation and cost effectiveness of the Digital Planimetric Compiler.

A possible minimum change consideration for the DPC would be to expand the subsystem's hardware and software so that each work station is assigned its own output magnetic tape. The physical tape unit may be assigned dynamically

at the start of job processing but in no case should more than one work station write output to a given tape unit at any one time. The software could also be expanded to permit multiple date files with each file representing one job to be stacked on the output tape. This would make it possible for one tape to contain several days' work for a given job prior to submission to the GIST preprocessor. Data identification on magnetic tape is currently made by the work station identification number. This job identification and accounting would be more meaningful and definitive if it were related to a job ID/charge-number/data scheme. These minimum changes would relieve or eliminate 6 of 10 deficiencies discussed above; there are, however, some additional modifications that should be considered.

A secondary change that could be made would be more difficult to implement with a smaller return on effort. This would involve either changing the DPC magnetic tape output format to the CALMA/NOVA notation or provide software to make the format transition. The existing PDP 11/20 would be the ideal place for the software to reside, but a detailed analysis of the new CPU time requirements (suggested minimum changes above) and the various techniques of implementing the new output format should be performed to determine the feasibility of this enhancement on the current PDP 11/20.

All the above alternatives add up to a significant reconfiguration of both the DPC software and hardware. In addition, the data base structure of the GIST system is not considered efficient for use in an interactive system. The question must be asked: Is this the point in time when the knowledge gained through the operation of the DPC subsystem, as well as the other known data entry devices, and the knowledge of the pros and cons of the various modules of the GIST subsystem should be combined and analyzed to determine the practicality of designing and implementing an interactive edit system on a dedicated midi/mini computer, using state-of-the-art devices for data capture and editing?

A discussion of some of the functional objectives and operational requirements of such a system is addressed later in this Appendix.

A.4.3.2 Comp-U-Grid (CEC) and CALMA (C-985). The CEC and the CALMA are both data entry systems used to convert analog source material (maps, plots, photo, etc.) to a digital form directly acceptable to the GIST system. While there are some hardware differences, these two systems are essentially the same. Each system has two strong points. First is that each output tape contains information from one work station (only one work station exists on each system), and the second is the output format is an absolute format (CALMA/NOVA) and is directly acceptable by the GIST system. Note that most of the faults cited for the DPC subsystem do not exist here, but there are potential improvements that should be addressed.

Each system, the CEC and the CALMA, is expensive for the rate of return in captured, verified digital data. This reflects two basic design features--one input work station per system, and a dependency on the batch operated UNIVAC 1108 GIST system to create a master file, verification plots, and to update (edit) master data files.

The existing hardware configuration in each system (excluding output tapes) is considered capable of supporting at least two input work stations with the appropriate software. The cost of an additional work station is a relatively small amount to increase the data capture rate and to insure that one inoperative work station does not remove a complete subsystem from the active production line.

The dependency on GIST to create a master data file, provide a verification plot, merge corrected digital data (results of an edit operation), and to add new digital data to a master file adds a large calendar delay time period to the processing sequence, reducing the effective capture rate of both systems. A side effect of the GIST turn around time is a decrease in the effective use of the cartographer due to fragmentation of his work on a given job. Several days may elapse waiting on UNIVAC 1108 turn around, and the cartographer may need a period for a reorientation to the job at hand.

A.5 SACARTS Expansion. The Topographic Center recognizes that SACARTS is in a state of dynamic evolution accommodating advancements of new and improved hardware systems. Associated with new equipment is acknowledged need to modify and improve the software techniques used to support and service these new devices. The most current expansion to SACARTS is the addition of the raster scanner/plotter, a device that should operate with a time/data capture rate that will prove cost effective when compared with the current methods of capturing digital data. A parallel development, which can function independently of the raster scanner/plotter or may be used to augment its operation, is the Digital Input/Output Display Equipment (DIODE). The DIODE is an interactive edit device which will permit the cartographer to make changes (corrections or additions) from an analog source directly to a digital data base (not the GIST data base).

A.5.1 Raster Scanner. The raster scanner/plotter serves two separate functions which can be used independently. The functions of the scanner would be performed in the following order:

First, the source document to be converted to a digital raster image is mounted on the scanner. The scan function is performed and a raster matrix is created.

Second, a lineal file using the new "MAPØ1" raster-to-linear software will be generated. The product lineal file is in a CALMA/NOVA format and is acceptable to GIST for subsequent edit procedures.

In addition to the "MAPØ1" software, RADC raster-to-lineal conversion software is being delivered to the DMA and installed on the DMAAC UNIVAC 1108 processor. This software, which could be adapted to GIST output, is used to illustrate the need for an edit capability. The remainder of this section discusses some of the considerations involved if the RADC-generated raster-to-linear software were used to create the lineal file from the raster matrix. This serves as background in discussing typical edit operations. The generated raster matrix is operated upon by a set of software used to first skeletonize (reduce) the volume of data and, second, to create a series of lineal features. These features are created and the data points linked based on the

distance relationship of one point to another. It should be noted that the output of this function is a group of features (point and linear) for which feature classification information is not and cannot be provided at this point. In addition, feature positional definition conflicts, such as a feature splitting into two legs forming an inverted Y, cannot be properly resolved by software. The feature definition and classification problems must be resolved by an edit operation.

Third, the lineal features are edited. Unfortunately, this is easier said than done. The output format generated by the RADC raster-to-linear software does not agree with any of the formats currently in use by SACARTS. None of the data entry devices (DPC, CEC, CALMA, or DIODE) can operate on this format, nor can the GIST system process the raster-to-linear output.

A reformat program would be required, but the more efficient approach in CPU time and therefore cost, would be to redefine the output format of the RADC raster-to-linear software to agree with one of the SACARTS operational data structures. Once the data is in the correct format, it will be possible to create a GIST master file with dummy header information attached to each feature. The edit operations required to accomplish feature classifications, resolve positional conflicts, and any other desired or necessary edit changes could be performed on any work station (DPC, CEC, CALMA, or DIODE) in the SACARTS.

A.5.2 Raster Plotter. The raster plotter has somewhat a different requirement. A software package is being provided to convert a raster matrix image to a string of commands (on/off, burn/non-burn) used to control the raster plotter and accomplish the generation of the output product. This requires the generation of the raster image from a linear file. Software is under development to make this conversion.

A.5.3 Digital Input/Output Display Equipment (DIODE). The DIODE is a high-powered interactive edit device. The options available for data selection and edit will make it a valuable addition to the SACARTS once it becomes operational. Its throughput rate will no doubt outstrip any work station currently in operation. Its interface into the SACARTS production system introduces two problems. First, the DIODE, accounting for the GIST deficiency in a file maintenance operation, uses a data structure that is more suited for file maintenance requirements. The result is that a GIST master file must be reformatted prior to DIODE processing and the DIODE digital data in turn must be converted to a GIST format after all processing has been completed. The second consideration is that the DIODE currently supports a single work station and regardless of its throughput rate is a prime candidate to become a bottleneck due to hardware malfunctions and the sheer volume of data to be processed.

A.6 Summary. Some of the operational and functional aspects of the SACARTS in its current configuration and with the enhancements of the DIODE and the raster scanner/plotter were discussed in this appendix. The objective of this discussion is to summarize some of the bottlenecks and the environmental conditions, in both hardware and software, causing these choke points.

The remainder of this section will briefly recap the major points discussed in the preceding portion of the SACARTS appendix and present several conclusions.

A.6.1 SACARTS Recap. GIST software operates on the UNIVAC 1108 in a batch environment. There are seven program modules which are the components of GIST. These seven modules comprise six independent software packages; modules 5 and 6 have been combined into one executable software package. Each package must be operated in the proper sequence and its results verified before the next software package in the production chain may be activated. Each software package requires human intervention by a cartographer for input data preparation and control card deck setup and by the UNIVAC 1108 operator for execution. The sequential nature and the need to verify the accuracy of the work performed combine to add a calendar delay time measured in days in accomplishing the UNIVAC 1108 processing. In addition, this sequence results in a proliferation of intermediate tapes. This increases the likelihood of confusion in tape handling and errors in reading or writing (parity) the data tapes.

The current GIST data structure does not lend itself to the rapid modification (edit) of a digital file. Digital data is organized in a GIST master file on a sector/page/node basis. Each node contains feature classification information, control data, and a series of encoded vectors which describe the convolutions of the line segment contained within the node. When performing an edit on an existing feature, two pieces of information are supplied--feature classification identification number range and the X, Y position of the work station stylus. A filtered serial search of each node in each page of the section selected by the X, Y position of the stylus is then required in an attempt to match the location identified by the cartographer via the stylus with a point found within a node that satisfies the feature classification identification range specified in the request.

The serial search of all nodes within the sector, even with the one element filter, is a sequential process that increases computer search time on the UNIVAC 1108. See section A.3 for a discussion of the GIST data structure with an example of a possible alternative approach to reduce or eliminate the time requirement for a sequential search of the GIST nodes in an attempt to locate a cartographer defined point.

Data capture devices (digitizing subsystems) are isolated from the GIST support software and cause a fragmentation of the work effort and cartographer's concentration. The three existing subsystems (DPC, CEC, and CALMA) are passive in nature. All actions performed at the work stations must be interpreted by the GIST UNIVAC 1108 software before a digital master file can be created or amended. The DPC further complicates the production operations because as many as five work stations may be intermixing data on one input tape that is in a format not acceptable for input to the GIST system. Several levels of DPC reconfiguration of hardware and software are discussed in section A.4; that section also addresses the CEC and CALMA digitizing subsystems. The data capture rate for the CEC and CALMA systems is restricted by virtue of only one work station associated with each system.

The raster scanner and its associated raster-to-linear software package (MAPØ1) will produce a lineal file containing linear and point features without any feature classification information. There is a need to edit the file and thereby classify features and resolve X, Y positional conflicts that occurred during the raster-to-linear conversion. This edit is not a trivial operation and with the current SACARTS equipment can be quite a time consuming procedure. However, the anticipated speed with which the entire process of scanning, linear conversion, and editing can be accomplished indicates a throughput rate that is likely to prove cost effective. Once the DIODE is fully operational and is inserted in the above procedure, a major improvement in cost effectiveness over the current SACARTS digitizing systems should be realized.

The DIODE operates on a data format that is foreign to GIST and all of the current SACARTS digitizing subsystems. The DIODE offers a dedicated minicomputer that allows the online interactive edit of a digital file. The edit functions will be augmented by a strong set of flexible options to support the cartographer when performing the necessary correction and addition to the digital file. All digital files must be reformatted on entry to and on exit from the DIODE system.

The DIODE, once fully operational, should provide a marked increase in editing throughput rate. It is, however, just one device and its ability to keep pace with the raster scanner's output is in doubt at this time. Documentation describing the current DIODE configuration at the Topographic Center indicates that it is capable of being expanded to two work stations. An evaluation of cost of expansion to a two work station DIODE system and an analysis of the benefit in the form of overall increase of throughput rate should be undertaken. Some key points to be considered in the benefit analysis are: degradation of system response time to the cartographer's commands, changes in bulk storage requirements, and variations in the mean-time-between-failure with the addition of the new equipment.

The raster plotter will generate a plot, regardless of the data volume or density, in approximately 15 minutes. A lineal file may not be plotted on the raster plotter. The lineal file must be converted to a format that is acceptable to the plotter. Traditionally, software used to convert the linear file to a raster file and then to a set of on/off burn/no-burn commands have been time consuming. The software is still under development, and final estimates of the throughput rate are not now available.

A.6.2 SACARTS Conclusions. The evolution of SACARTS to its present status has undergone an extended period of growth as additional requirements and capabilities were introduced. Its transition to a high-volume production support system has involved a concerted and committed effort on the part of personnel in the Department of Cartography to fully exploit the system capabilities to meet these production needs. Many of the limitations of the current system have already been recognized by these personnel and have resulted in recommendations for system change or enhancement. Introduction of new equipment such as the DIODE and the Scanner/Plotter together with expanded system software capabilities reflect further evolutionary growth of system capacity

and flexibility. This growth also indicates an increasing trend toward interactive operations with more direct participation by the cartographer functioning in an interactive mode. This analysis has therefore concentrated on examining areas which the study team feels will provide the greatest benefit to the cartographer in expanding the throughput capacity of the SACARTS system. Within this context the following conclusions were drawn:

- Augmented digitizing systems (CEC and CALMA) could increase the data/edit capture rate.

- An augmented DPC would relieve a burdensome data preprocessing/reformat problem.

- Augmented digitizing systems will not resolve:
  - -- The UNIVAC 1108 processing work load for data entry, edit, or feature search time.
  - -- The calendar time required to sequence GIST software modules 1 through 4.
  - -- The volume of intermediate magnetic tapes to be handled, processed, and journalled.

- GIST modules 5/6 and 7 do not suffer in operational efficiency due to the batch environment of the GIST/UNIVAC 1108 structure.

- GIST modules 1 through 4 require frequent human intervention by both the cartographer and the computer operator.

A.6.3 Recommendations. As a result of this analysis, the following recommendations are made:

- A dynamic online interactive data capture/edit system encompassing the functions of GIST modules 1 through 4 should be considered. A functional description of such a system is provided in Appendix B. Some desirable objectives are:
  - -- Dedicated minicomputer with multiple smart work stations
  - -- Full interactive edit capacity operating directly on the digital data base
  - -- Localized quick-look proof plot capability
  - -- Cartesian coordinate conversion to geographic coordinates and vice versa
  - -- Reformat to DMA standard for planimetric data format
  - -- Mainframe CPU foreground/background operation

- The system selected to perform the interactive edit function in support of SACARTS should be of a design that can be easily adapted to meet interactive edit requirements of the Field Offices.

APPENDIX B

INTERACTIVE EDIT STATION FUNCTIONAL DESCRIPTION

The SACARTS evaluation identified a major problem area, the calendar delay time required to generate an error-free digital data file. The delay time is the result of the batch mode environment of GIST modules 1 through 4.

The delay arises from the sequence in which every action performed by the cartographer at a work station must be recorded on magnetic tape and physically moved to the UNIVAC 1108 computer area. One way to offset this calendar delay time is to use a dedicated interactive minicomputer network that affords the cartographer the ability to create a data file in the proper format and allows the cartographer to interactively modify or edit that data file. A detailed discussion of such a system is outside the scope of the SACARTS evaluation task, but, the tremendous impact such a system would have on the data capture/edit procedures demands that a few words must be included. Figure B-1 illustrates a typical interactive data capture/edit system. The remainder of this section will address some of the considerations associated with the smart work stations, communication interface, and the mainframe CPU of such a system. The system is described in terms of functions and does not describe any specific system known to exist currently.

B.1 Smart Work Station. The incorporation of a minicomputer at the work station coupled with a prudent system design makes it possible for each work station to create and edit digital data as if it were the only station operating. All data captured at a work station is in the proper format; the need for preprocessing/reformatting is eliminated.

All data capture/edit operations are performed within the work station (CPU). The cartographer at the work station may direct the mainframe CPU to perform operations on a data file such as save file to magnetic tape, transform to a geo-coordinate system, etc. (see Background Functions below).

B.2 Desirable Work Station Functions. Work station capabilities should include the ability to perform the following functions:

- Select the mode of operation. Reducing work station software to categories (or modes) of operation (digitize data, edit, etc.) makes it possible to decrease the volume of software required in the work station CPU at any one time. This allows for a more efficient and economic use of core. Some of the operational modes that should be considered are:
  - -- Job log on/off, fetch job accounting software and initialize/report the job statistics associated with this job
  - -- File header define, name the file, and supply cartographic parameters associated with the new data file
  - -- Source document registration on digitizing table
  - -- Define feature classification, classify or reclassify a feature

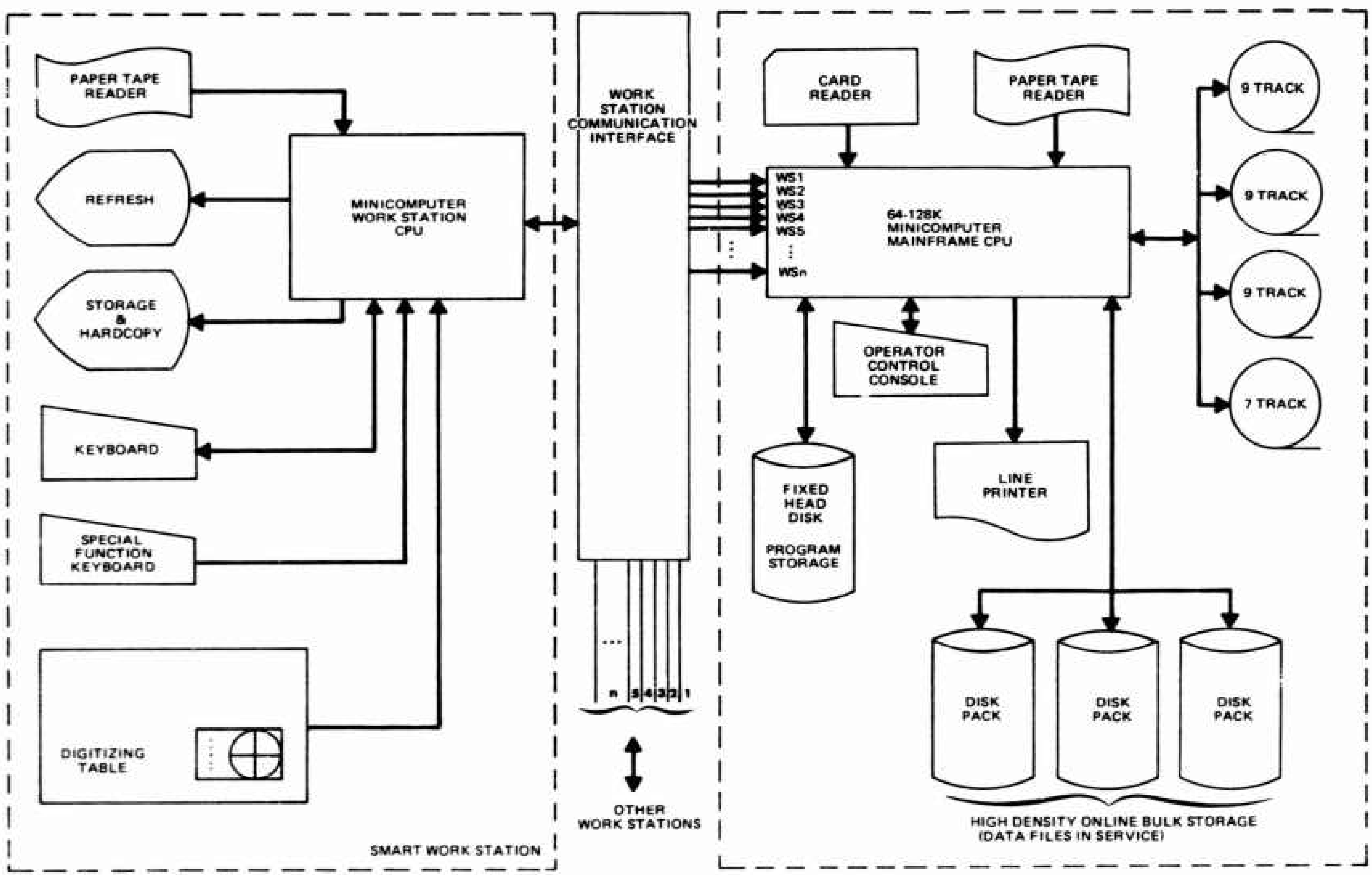


Figure B-1. A Typical Cartographic System

-- Digitize a feature, interpret the movement of the stylus, and create a digital feature. Both trace and point-to-point modes are honored in any combination.
-- Edit a feature, an additional modifier is required to select the desired edit option (see Edit digital data below)
-- Remote Job Entry is used to specify a job(s) to be performed on and by the mainframe CPU. The jobs that may be executed are listed in the Mainframe CPU discussion that follows.

- Define job accounting information for journaling

- Define file header data: lat/long of reference corner and control points, projection type, etc. (projection type required only if a geo-transformation is anticipated)

- Register source document to digitizing table, generate parameters to compensate for rotation, translation, and skew of source document placement on digitizing table

- Enter feature classification information, permit multiple feature headers to be defined when processing in either the digitize or edit mode

- Digitize data in any combination of trace and point-to-point modes. If multiple feature headers have been defined, create duplicate features.

- Edit digital data, the following functions should be included:
  -- Delete a feature from the file
  -- Delete a segment of a feature, possibly resulting in two features
  -- Change classification information
  -- Feature end segment modify, manually change the beginning or end point of a feature
  -- Modify the mid-segment of a feature
  -- Split a feature into two or more features and change classification information if desired
  -- Join two features

B.3 Work Station Hardware. The following hardware illustrated in Figure B-1 would be required for the work station:

- Minicomputer containing 16K of core controls all phases of the operations at work station as well as the asynchronous communications with the mainframe CPU via the communication interface device.

- Paper tape reader serves as the input device for work station diagnostic software.

- Refresh CRT presents a dynamic presentation of a feature on which the cartographer is performing an edit.

- Storage CRT with an optional hard copy capability will serve several functions. Menus guiding or prompting the cartographer in a sequence of manual operations required to perform a task will be presented; messages from the work station CPU (i.e., job complete, job abort, etc.) will be displayed, and the storage CRT provides an economic means by which a family of features around a designated point can be displayed. The economy is realized through the saving of core required by a refresh CRT that is not required by a storage CRT.

- Alphanumeric keyboard is the command input device which permits the cartographer to direct operations to be performed by the work station CPU. Individual commands and alphanumeric text may be entered.

- Special function keyboard will augment the alphanumeric keyboard. Most commonly used commands will be selectable via one button depression reducing the number of physical actions required on the part of the cartographer.

- Digitizing table is a data entry (capture/edit) device. It provides the cartographer with the ability to classify, create digital data, and to modify and edit existing digital data.

B.4 Communication Interface. This is an asynchronous device used to control the transfer of data between each work station and the mainframe CPU. The asynchronous aspect of the device ameliorates the timing constraints that exist in the mainframe CPU. Each work station is connected through dedicated lines through the Communication Interface to the mainframe CPU insuring immediate access to the mainframe CPU. The data transfers may be either serial or parallel word transfers. Each word transfer should be controlled by a ready/resume signal exchange between the two CPU's to eliminate the possible loss of data due to an overwrite during the data exchange.

B.5 Mainframe CPU. This CPU serves two functions: full and immediate support of the work being performed at the work station, and honoring job requests initiated at the work station of a quasi off-line nature, i.e., file save to magnetic tape, generate a plot tape, etc. (see Background Functions below). To support these functions the CPU will operate in a foreground/background environment.

Requests to retrieve work station programs and to store/retrieve digital data are foreground requests and are of the highest priority. They will be honored immediately on receipt in the mainframe CPU. Any background processing will be interrupted to permit a foreground request to be honored.

The software provided with this CPU includes a complete data base management system to support the storage and retrieval of digital data created/edited at the work station and stored on the bulk storage disk included in the mainframe CPU.

B.6 Background Functions. Some desirable background functions are the abilities to:

- Save a digital file on magnetic tape for archival storage and/or input to GIST modules 5/6
- Restore a digital file from magnetic tape to permit a new round of editing
- Transform a two dimensional Cartesian digital file to a spherical geographic coordinate file
- Transform the geographic coordinate file to a Cartesian coordinate file
- Polygon extraction, used to create a new file that will contain the digital data inside or outside, as specified, of the cartographer defined polygon
- Proof plot generate
- Merge two files into a new file
- Filter a file, extract selected features from a parent file and create a new file. The cartographer will dynamically set the feature selection parameters directing the filter operation.
- Panel two files and create a third file. Two files geographically adjacent to each other are candidates for a panel (not merge) operation. When panelling a traveling feature, such as a road that crosses from one file into the other, the result would be one continuous feature in the product file. In a merge function the product file would preserve the two original features as separate entities.

B.7 Mainframe CPU Hardware. The mainframe CPU hardware would function as follows:

- Mainframe CPU will provide all support software required by the various work stations, and the data base management system to store and retrieve digital data from bulk storage; control data transfer to and from magnetic tape; provide work station software on request; and perform special software functions (see Background Functions above).
- Card reader--data input device primarily used for software development (optional).
- Paper tape reader--input device for mainframe CPU diagnostic software.

- Operator control console--operator command input device.

- Line printer--used as an output report device.

- Fixed head disk--central storage device for all work station CPU and mainframe CPU software. Speed of access makes the fixed head disk more desirable than a removable disk pack unit.

- Bulk storage--online bulk storage is used to store the active digital files (in use at the work station). All access to these files is made through the data base management system of the mainframe CPU software.

- Magnetic tapes--serves several functions: archival storage for error free files, plot tapes for data verification, data input device for archival error free files or for data files generated by other agencies, and data output device to allow data to be distributed to other agencies.

BIBLIOGRAPHY

Anderson, J. P., Multics Evaluation--Final Report, Report #ESD-TR-73-276, October 1973.

Baldauf, David L., "Experiences with an Operational Associative Processor," G. Goos and J. Hartmanis (eds.), Lecture Notes in Computer Science, Parallel Processing, Proceedings of the Sagamore Computer Conference, August 1974, pp. 270-271.

Bauer, Lyal H., "Implementation of Data Manipulating Functions on the STARAN Associative Processor," G. Goos and J. Hartmanis (eds.), Lecture Notes in Computer Science, Parallel Processing, Proceedings of the Sagamore Computer Conference, August 1974, pp. 209-227.

Branstad, D. K., "Security Aspects at Computer Networks," AIAA Computer Network Conference, Huntsville, Alabama, April 1973.

Cullinane, J., et al., Commercial Data Management Processor Study (Babbage), Cullinane Corporation, Wellesley, Massachusetts, December 1975.

Curran, L., "Air Force 'Kernel' Attains Computer Security Using Existing Technology," Electronics, Volume 49, September 30, 1976, p. 59.

DeFiore, D. R., et al., "Associative Techniques in the Solution of Data Management Problems," Proceedings ACM National Conference, 1971, ACM, pp. 28-36.

Flynn, M. J., "Some Computer Organizations and their Effectiveness," IEEE Trans. Computers, September 1972, pp. 248-960.

Gillotte, M., N. Andres, and M. Birdsall, ACS Data Bank Development--Final Report Contract F30602-75-C-0008 Pattern Analysis and Recognition Corporation, Rome, New York, July 1976. RADC-TR-76-372. AD#B016452L.

Goodyear, "STARAN APPLE Programming Manual," Document GER-1563B, September 1974.

Gudes, E., and H. S. Koch, "The Application of Cryptography for Data Base Security," Proceeding of NCC, June 1976, p. 97.

Heinrich, F. R., and D. J. Kaufman, "A Centralized Approach to Computer Network Security," Proceedings of NCC, June, 1976, p. 85.

Higbie, L. C., "Supercomputer Architecture," Computer, December 1973, pp. 48-58.

Ingemarsson, I., "Analysis of Secret Functions with Application to Computer Cryptography," Proceedings of NCC, June, 1976, page 125.

Johnson, M. D. and D. C. Gunderson, "An Associative Data Acquisition System," Proceedings 1970 International Telemetry Conference, April 1970, pp. 107-115.

Kaufman, D. J., "A Distributed Approach to Computer Network Security," System Development Corporation, SP-3848, May 31, 1976.

Lampson, B. W., "Dynamic Protection Structure," Fall Joint Computer Conference, 1967.

Lowenthal, E. I., The Back-End Computer, MRI Systems Corporation, April 1976.

Martin, J., Security, Accuracy, and Privacy in Computer Systems, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1973.

Moulder, R., "An Implementation of a Data Management System on an Associative Processor," NCC, 1973.

"Privacy and Security in Computer Systems," NBS Special Publication #404, Proceedings of a Conference held at NBS, March 1974.

Radosevic, R. G., et al., "Associative Array Processing of Raster Scanned Data for Automated Cartography," Final Report ETL-0046, March 1976.

Ruben, S., et al., "Application of a Parallel Processing Computer in LACIE," paper presented at International Conference on Parallel Processing at Waldenwoods, Michigan, August 1976.

Schmitz, H. G., et al., "ABMDA Prototype Bulk Filter Development Concept Definition Phase," Final Report, Contract No. DAH60-72-C-0050, NTIS, April 1972.

Shore, J. E., "Second Thoughts on Parallel Processing," Computing and Electrical Engineering, Volume 1, Pergannon Press, Oxford, England, 1973, pp. 95-109.

Thurber, K. J., "An Associative Processor for Air Traffic Control," Proceedings AFIPS Spring Joint Computer Conferences, 1971, pp. 49-59.

Wald, L. D., "Integrated Voice/Data Compression and Multiplexing using Associative Processing," Proceedings AFIPS, 1974 NCC, May 1974, pp. 133-138.

Wald, L. D. and G. A. Anderson, "Associative Memory for Multiprocessor Control," Final Report NAS, 12-2087, September 1971.

METRIC SYSTEM

BASE UNITS:

Quantity	Unit	SI Symbol	Formula
length	metre	m	...
mass	kilogram	kg	...
time	second	s	...
electric current	ampere	A	...
thermodynamic temperature	kelvin	K	...
amount of substance	mole	mol	...
luminous intensity	candela	cd	...

SUPPLEMENTARY UNITS:

Quantity	Unit	SI Symbol	Formula
plane angle	radian	rad	...
solid angle	steradian	sr	...

DERIVED UNITS:

Quantity	Unit	SI Symbol	Formula
Acceleration	metre per second squared	...	m/s
activity (of a radioactive source)	disintegration per second	...	(disintegration)/s
angular acceleration	radian per second squared	...	rad/s
angular velocity	radian per second	...	rad/s
area	square metre	...	m
density	kilogram per cubic metre	...	kg/m
electric capacitance	farad	F	A·s/V
electrical conductance	siemens	S	A/V
electric field strength	volt per metre	...	V/m
electric inductance	henry	H	V·s/A
electric potential difference	volt	V	W/A
electric resistance	ohm		V/A
electromotive force	volt	V	W/A
energy	joule	J	N·m
entropy	joule per kelvin	...	J/K
force	newton	N	kg·m/s
frequency	hertz	Hz	(cycle)/s
illuminance	lux	lx	lm/m
luminance	candela per square metre	...	cd/m
luminous flux	lumen	lm	cd·sr
magnetic field strength	ampere per metre	...	A/m
magnetic flux	weber	Wb	V·s
magnetic flux density	tesla	T	Wb/m
magnetomotive force	ampere	A	...
power	watt	W	J/s
pressure	pascal	Pa	N/m
quantity of electricity	coulomb	C	A·s
quantity of heat	joule	J	N·m
radiant intensity	watt per steradian	...	W/sr
specific heat	joule per kilogram-kelvin	...	J/kg·K
stress	pascal	Pa	N/m
thermal conductivity	watt per metre-kelvin	...	W/m·K
velocity	metre per second	...	m/s
viscosity, dynamic	pascal-second	...	Pa·s
viscosity, kinematic	square metre per second	...	m/s
voltage	volt	V	W/A
volume	cubic metre	...	m
wavenumber	reciprocal metre	...	(wave)/m
work	joule	J	N·m

SI PREFIXES:

Multiplication Factors	Prefix	SI Symbol
1 000 000 000 000 = 10^{12}	tera	T
1 000 000 000 = 10^{9}	giga	G
1 000 000 = 10^{6}	mega	M
1 000 = 10^{3}	kilo	k
100 = 10^{2}	hecto*	h
10 = 10^{1}	deka*	da
0.1 = 10^{-1}	deci*	d
0.01 = 10^{-2}	centi*	c
0.001 = 10^{-3}	milli	m
0.000 001 = 10^{-6}	micro	μ
0.000 000 001 = 10^{-9}	nano	n
0.000 000 000 001 = 10^{-12}	pico	p
0.000 000 000 000 001 = 10^{-15}	femto	f
0.000 000 000 000 000 001 = 10^{-18}	atto	a

* To be avoided where possible.

MISSION
of
Rome Air Development Center

RADC plans and conducts research, exploratory and advanced development programs in command, control, and communications (C^3) activities, and in the C^3 areas of information sciences and intelligence. The principal technical mission areas are communications, electromagnetic guidance and control, surveillance of ground and aerospace objects, intelligence data collection and handling, information system technology, ionospheric propagation, solid state sciences, microwave physics and electronic reliability, maintainability and compatibility.

AMERICAN REVOLUTION BICENTENNIAL 1776-1976